A Geometric Approach to Dynamical System: Global Analysis for Non-Convex Optimization

Ji Xu

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

ABSTRACT

A Geometric Approach to Dynamical System: Global Analysis for Non-Convex Optimization

Ji Xu

Non-convex optimization often plays an important role in many machine learning problems. Study the existing algorithms that aim to solve the non-convex optimization problems can help us understand the optimization problem itself and may shed light on developing more effective algorithms or methods. In this thesis, we study two popular non-convex optimization problems along with two popular algorithms.

The first pair is maximum likelihood estimation with expectation maximization algorithm. Expectation Maximization (EM) is among the most popular algorithms for estimating parameters of statistical models. However, EM, which is an iterative algorithm based on the maximum likelihood principle, is generally only guaranteed to find stationary points of the likelihood objective, and these points may be far from any maximizer. We address this disconnect between the statistical principles behind EM and its algorithmic properties. Specifically, we provide a global analysis of EM for specific models in which the observations comprise an i.i.d. sample from a mixture of two Gaussians. This is achieved by (i) studying the sequence of parameters from idealized execution of EM in the infinite sample limit, and fully characterizing the limit points of the sequence in terms of the initial parameters; and then (ii) based on this convergence analysis, establishing statistical consistency (or lack thereof) for the actual sequence of parameters produced by EM.

The second pair is phase retrieval problem with approximate message passing algorithm. Specifically, we consider an ℓ_2-regularized non-convex optimization problem for recovering signals from their noisy phaseless observations. We design and study

the performance of a message passing algorithm that aims to solve this optimization problem. We consider the asymptotic setting $m, n \to \infty$, $m/n \to \delta$ and obtain sharp performance bounds, where m is the number of measurements and n is the signal dimension. We show that for complex signals the algorithm can perform accurate recovery with only $m = \frac{64}{\pi^2} - 4 \approx 2.5n$ measurements. The sharp analyses in this paper enable us to compare the performance of our method with other phase recovery schemes.

Finally, the convergence analysis of the iterative algorithms are done by a geometric approach to dynamical systems. By analyzing the movements from iteration to iteration, we provide a general tool that can show global convergence for many two dimensional dynamical systems. We hope this can shed light on convergence analysis for general dynamical systems.

Table of Contents

List of Figures

Acknowledgments

I would like to express my most sincere gratitude to my advisors Professor Arian Maleki and Professor Daniel Hsu. Both of them have provided me with tremendous guidance throughout my studies at Columbia. I met Arian first when I took his courses, and later I was his research assistant during my master's program. He is extremely patient and skillful in instructing and inspiring his students. In fact, without the inspiration of his enthusiasm for exploring and working on research, I would not further pursue my Ph.D. degree, which would be a life changing event. I did not think that l could be any luckier after having Arian being my mentor. It turns out that I was wrong and Daniel is also an excellent mentor that every Ph.D. student would have hoped for. Besides his brilliance on solving the research problems, he also has an extraordinary vision of raising good research problems. He always encourages me to look at my research in a larger scope and find problems that potentially have larger impact. It is the consistent and generous guidance from both Arian and Daniel that I have developed my own research interests, one of which is non-convex optimization, the topic of my thesis. It is impossible for me to complete my degree and have such an enjoyful Ph.D. life without Arian and Daniel.

I would also like to thank all of my committee members. I was in Professor Alexandr Andoni's advanced algorithms course in my second year. Because of his excellent teaching, I have learned a lot about many classical algorithmic ideas such as sketching and locality-sensitive hashing (LSH). Professor John Paisley was on my oral

exam and both Alex and John attended my thesis proposal. I want to thank them for their thoughtful questions and insightful comments that lead to some promising future directions of this thesis. I also want to thank Professor Nakul Verma, who is my last committee member. We met at Daniel's grouping meeting. His participation always adds depth to the discussions and benefits the entire group including me. We also met at the poster section in the last year NeurIPS conference. Again, I benefit a lot from our discussion about the research presented at the conference.

I would like to thank the faculties members and staff of the Department of Computer Science and also Department of Statistics for their great support. First, I would like to thank my co-author Dr. Junjie Ma. Because of our successful and enjoyable collaborations, we have finished several research projects together, part of which has become a major chapter of my thesis. Then, I would like to thank Professor Victor de la Pena for his help in my Ph.D. application process. Also, his teaching of my probability course has sharpened my knowledge about probability inequalities. Then I would like to thank Jessica Rosa, Cynthia Meekins, Dood Kalicharan and Anthony Cruz for their great support for all Ph.D. students in Computer Science and Statistics department. Because of their excellent administrative work including organizing seminars, happy hours and helping with documents, I can widen my knowledge from numerous talks, have fun with friends and pay all my attention to the research without any concerns of paperwork.

I would like to thank my friends Dr. Da Tang, Dr. Kevin Shi, Peilin Zhong, Rishabh Dudeja, Wenda Zhou, Kiran Vodrahalli, Giannis Karamanolakis, among others, for the motivative discussions on research. Besides research discussion, I would also like to thank my friends, Dr. Yang Kang, Dr. Jing Wu, Dr. Xiaopei Zhang, Dr. Da Tang, Dr. Xun Sun, Dr. Yuanjun Gao, Dr. Xiaowei Tan, Peilin Zhong, Chang Xiao, Fengpei Li, Weizhuo Sun, among others, for their kind help and support. Because

of them, I had wonderful trips in many places in and outside America, and all those happy hours and game nights in the city. They made my life enjoyable and colorful.

Finally, I would like to thank my parents Minli Wang and Zhiqiang Xu for their unconditional love and encouraging support. Undoubtedly, I could not make this far without them. I also appreciate the support from my girlfriend Yilin Sun.

To My Family.

Chapter 1

Introduction and Background

1.1 Introduction

Optimization often plays an important role in many machine learning problems, from the simplest linear regression model to more complicated deep learning models. Depends on the objective function, the optimization problems are divided into two categories: convex and non-convex optimization. Comparing to convex optimization, a non-convex optimization can capture the learning and prediction problems more accurately, but in general, such optimization could be NP-hard to solve. To address the intractability issue, a common approach, known as convex relaxation, is to transform the non-convex objective to a convex one and mainly focus one the later convex optimization problem. Although this approach is often convenient in derivation, it loses the immense modeling power given by the non-convex optimization. Hence, another approach, perhaps a more appropriate one, is to analyze the non-convex problems directly. Further, despite of the possibility of spurious local optima, simple iterative methods such as gradient descent have been remarkably successful to solve non-convex optimizations in practice [Jain and Kar, 2017; Chi *et al.*, 2018]. However, the theoret-

ical footings for the good practical performance, in contrast, had been largely lacking until recently [Candès and Recht, 2009; Davenport and Romberg, 2016; Chen and Chi, 2018; Shechtman *et al.*, 2015]. They demonstrated that global optimal can be found in some non-convex problems that arise in natural settings such as matrix completion and phase retrieval. Their success encourage us to analyze more complicated non-convex problems in machine learning and signal processing. More specifically, the goal is to answer the following questions for a given non-convex optimization problem

Q.1 Are all local optima are global optima?

Q.2 If local optimum exists, when a common iterative algorithm (e.g. gradient decent) can find the global optimum? What initialization or regularization is required for the algorithm to find the global optimum?

The answers to these questions can provide us a deeper understanding of the non-convex optimizations that appear in practice and may shed light on developing more effective algorithms or methods to solve these problems. The problems we consider in this thesis are mainly based on two areas: maximum likelihood estimation and low rank matrix estimation. The related work in these areas as well as other interesting areas are presented in the following subsections.

1.2 Maximum Likelihood Estimation

Since Fisher's 1922 paper Fisher [1922], maximum likelihood estimators (MLE) have become one of the most popular tools in many areas of science and engineering. The asymptotic consistency and optimality of MLEs have provided users with the confidence that, at least in some sense, there is no better way to estimate parameters for many standard statistical models. Despite its appealing properties, computing

the MLE is often intractable. Indeed, this is the case for many *latent variable models* $\{f(\mathcal{Y}, \boldsymbol{z}; \boldsymbol{\theta})\}$, where the *latent variables* $\boldsymbol{z}$ are not observed. For each setting of the parameters $\boldsymbol{\theta}$, the marginal distribution of the observed data $\mathcal{Y}$ is (for discrete $\boldsymbol{z}$)

$$f(\mathcal{Y}; \boldsymbol{\theta}) \;=\; \sum_{\boldsymbol{z}} f(\mathcal{Y}, \boldsymbol{z}; \boldsymbol{\theta})\,.$$

It is this marginalization over latent variables that typically causes the computational difficulty. Furthermore, many algorithms based on the MLE principle are only known to find stationary points of the likelihood objective (e.g., local maxima), and these points are not necessarily the MLE.

Among the algorithms mentioned above, Expectation Maximization (EM) has attracted more attention for the simplicity of its iterations, and its good performance in practice [Dempster *et al.*, 1977; Redner and Walker, 1984]. EM is an iterative algorithm for climbing the likelihood objective starting from an initial setting of the parameters $\hat{\boldsymbol{\theta}}^{\langle 0 \rangle}$. In iteration t, EM performs the following steps:

$$\text{E-step:} \qquad \hat{Q}(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{\langle t \rangle}) \;\triangleq\; \sum_{\boldsymbol{z}} f(\boldsymbol{z} \mid \mathcal{Y}; \hat{\boldsymbol{\theta}}^{\langle t \rangle}) \log f(\mathcal{Y}, \boldsymbol{z}; \boldsymbol{\theta})\,, \tag{1.1}$$

$$\text{M-step:} \qquad \hat{\boldsymbol{\theta}}^{\langle t+1 \rangle} \;\triangleq\; \arg\max_{\boldsymbol{\theta}} \hat{Q}(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{\langle t \rangle})\,, \tag{1.2}$$

In many applications, each step is intuitive and can be performed very efficiently.

Despite the popularity of EM, as well as the numerous theoretical studies of its behavior, many important questions about its performance—such as its convergence rate and accuracy—have remained unanswered. In Chapter 2, we address these questions for some specific Gaussian mixture models in which the observation $\mathcal{Y}$ is an i.i.d. sample from a mixture of Gaussians.

1.2.1 Background

The EM algorithm was formally introduced by Dempster *et al.* [1977] as a general iterative method for computing parameter estimates from incomplete data. Although EM is billed as a procedure for maximum likelihood estimation, it is known that with certain initializations, the final parameters returned by EM may be far from the MLE, both in parameter distance and in log-likelihood value [Wu, 1983]. Several works characterize local convergence of EM to stationary points of the log-likelihood objective under certain regularity conditions [Wu, 1983; Tseng, 2004; Chrétien and Hero, 2008]. However, these analyses do not distinguish between global maximizers and other stationary points (except, e.g., when the likelihood function is unimodal). Thus, as an optimization algorithm for maximizing the log-likelihood objective, the "worst-case" performance of EM is somewhat discouraging.

For a more optimistic perspective on EM, one may consider a "best-case" analysis, where (i) the data are an iid sample from a distribution in the given model, (ii) the sample size is sufficiently large, and (iii) the starting point for EM is sufficiently close to the parameters of the data generating distribution. Conditions (i) and (ii) are ubiquitous in (asymptotic) statistical analyses, and (iii) is a generous assumption that may be satisfied in certain cases. Redner and Walker [1984] show that in such a favorable scenario, EM converges to the MLE almost surely for a broad class of mixture models. Moreover, recent work of Balakrishnan *et al.* [2017] gives non-asymptotic convergence guarantees in certain models; importantly, these results permit one to quantify the accuracy of a pilot estimator required to effectively initialize EM. Thus, EM may be used in a tractable two-stage estimation procedures given a first-stage pilot estimator that can be efficiently computed.

Indeed, for the special case of Gaussian mixtures, researchers in theoretical com-

puter science and machine learning have developed efficient algorithms that deliver the highly accurate parameter estimates under appropriate conditions. Several of these algorithms, starting with that of Dasgupta [1999], assume that the means of the mixture components are *well-separated*—roughly at distance either d^α or k^β for some $\alpha, \beta > 0$ for a mixture of k Gaussians in $\mathbb{R}^d$ [Dasgupta, 1999; Arora and Kannan, 2005; Dasgupta and Schulman, 2007; Vempala and Wang, 2004; Kannan *et al.*, 2008; Achlioptas and McSherry, 2005; Chaudhuri and Rao, 2008; Brubaker and Vempala, 2008; Chaudhuri *et al.*, 2009b]. More recent work employs the method-of-moments, which permit the means of the mixture components to be arbitrarily close, provided that the sample size is sufficiently large [Kalai *et al.*, 2010; Belkin and Sinha, 2010; Moitra and Valiant, 2010; Hsu and Kakade, 2013; Hardt and Price, 2015]. In particular, Hardt and Price [2015] characterize the information-theoretic limits of parameter estimation for mixtures of two Gaussians, and that they are achieved by a variant of the original method-of-moments of Pearson [1894].

Most relevant to this thesis are works that specifically analyze EM (or variants thereof) for Gaussian mixture models, especially when the mixture components are well-separated. Xu and Jordan [1996] show favorable convergence properties (akin to super-linear convergence near the MLE) for well-separated mixtures. In a related but different vein, Dasgupta and Schulman [2007] analyze a variant of EM with a particular initialization scheme, and proves fast convergence to the true parameters, again for well-separated mixtures in high-dimensions. For mixtures of two Gaussians, it is possible to exploit symmetries to get sharper analyses. Indeed, Chaudhuri *et al.* [2009a] uses these symmetries to prove that a variant of Lloyd's algorithm [MacQueen, 1967; Lloyd, 1982] (which may be regarded as a hard-assignment version of EM) very quickly converges to the subspace spanned by the two mixture component means, without any separation assumption. Lastly, let us consider a specific case where the observa-

tion $\mathcal{Y}$ is an i.i.d. sample from the mixture distribution $0.5N(-\boldsymbol{\theta}^\star, \boldsymbol{\Sigma}) + 0.5N(\boldsymbol{\theta}^\star, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ is a known covariance matrix in $\mathbb{R}^d$, and $\boldsymbol{\theta}^\star$ is the unknown parameter of interest. Balakrishnan *et al.* [2017]; Daskalakis *et al.* [2017] proves linear convergence of EM for this specific case.

1.3 Phase Retrieval

Phase retrieval refers to the task of recovering a signal $\boldsymbol{x}^\star \in \mathbb{C}^{n\times 1}$ from its m phaseless linear measurements:

$$y_a = \left| \sum_{i=1}^{n} A_{ai} x_i^\star \right| + w_a, \quad a = 1, 2, \ldots, m, \tag{1.3}$$

where $x_i^\star$ is the ith component of $\boldsymbol{x}^\star$ and $w_a \sim \mathcal{CN}(0, \sigma_w^2)$ a Gaussian noise. The recent surge of interest has led to a better understanding of the theoretical aspects of this problem [Candès *et al.*, 2013; Netrapalli *et al.*, 2013; Eldar and Mendelson, 2014; Candès *et al.*, 2015; Chen and Candès, 2017; Wang *et al.*, 2016; Zhang and Liang, 2016; Goldstein and Studer, 2016; Bahmani and Romberg, 2016; Cai *et al.*, 2016; Sun *et al.*, 2016; Soltanolkotabi, 2017; Duchi and Ruan, 2017; Lu and Li, 2017; Davis *et al.*, 2017; Soltanolkotabi, 2017; Tan and Vershynin, 2017; Jeong and Güntürk, 2017; Zeng and So, 2017; Mondelli and Montanari, 2017; Dhifallah and Lu, 2017; Dhifallah *et al.*, 2017; Abbasi *et al.*, 2017; Qu *et al.*, 2017]. Thanks to such research we now have access to several algorithms, inspired by different ideas, that are theoretically guaranteed to recover $\boldsymbol{x}^\star$ exactly in the noiseless setting. Despite all this progress, there is still a gap between the theoretical understanding of the recovery algorithms and what practitioners would like to know. For instance, for many algorithms, including Wirtinger flow [Candès *et al.*, 2015; Chen and Candès, 2017] and amplitude

flow [Wang *et al.*, 2016; Zhang and Liang, 2016], the exact recovery is guaranteed with either $cn \log n$ or cn measurements, where c is often a fixed but large constant that does not depend on n. In both cases, it is often claimed that the large value of c or the existence of $\log n$ is an artifact of the proving technique and the algorithm is expected to work with cn for a reasonably small value of c. Such claims have left many users wondering which algorithm should we use? Since the theoretical analyses are not sharp, they do not shed any light on the relative performance of different algorithms. Researchers have developed certain intuition based on a combination of theoretical and empirical results, to give heuristic answers to these questions. However, as demonstrated in a series of papers in the context of compressed sensing, such folklores are sometimes inaccurate [Zheng *et al.*, 2017]. In the light of this, we would like to further ask the following questions despite that we already have some answers to the question Q2 we have raised above.

Q.3 How does the constant δ affect the landscape of the optimization and the dynamics of the algorithm?

Q.4 Can we determine the constant δ for an algorithm?

Q.5 What is the impact of initialization schemes, such as spectral initialization?

To address these questions, several researchers have adopted the asymptotic framework $m, n \to \infty$, $m/n \to \delta$, and provided sharp analyses for the performance of several algorithms [Dhifallah and Lu, 2017; Dhifallah *et al.*, 2017; Abbasi *et al.*, 2017]. This line of work studies recovery algorithms that are based on convex optimization. In this thesis, we adopt the same asymptotic framework and study the following popular non-convex problem, known as amplitude-based optimization [Zhang and Liang,

2016; Wang *et al.*, 2016]:

$$\min_{\boldsymbol{x}} \quad \sum_{a=1}^{m} (y_a - |(\boldsymbol{A}\boldsymbol{x})_a|)^2 + \frac{\mu_k}{2}\|\boldsymbol{x}\|_2^2. \tag{1.4}$$

where $(\boldsymbol{A}\boldsymbol{x})_a$ denotes the a-th entry of $\boldsymbol{A}\boldsymbol{x}$. Note that compared to the optimization problem discussed in Zhang and Liang [2016]; Wang *et al.* [2016], Equation (1.4) has an extra ℓ_2-regularizer. Regularization is known to reduce the variance of an estimator and hence is expected to be useful when $\boldsymbol{w} \neq \boldsymbol{0}$. However, as we will try to clarify later in Chapter 3, since the loss function $\sum_{a=1}^{m} (y_a - |(\boldsymbol{A}\boldsymbol{x})_a|)^2$ is non-convex, regularization can help the iterative algorithm that aims to solve Equation (1.4) even in the noiseless settings.

Since Equation (1.4) is a non-convex problem, the algorithm to solve it matters. In this paper, we study a message passing algorithm that aims to solve Equation (1.4). As a result of our studies we

1. present sharp characterization of the mean square error (even the constants are sharp) in both noiseless and noisy settings.

2. present a quantitative characterization of the gain initialization and regularization can offer to our algorithms.

Furthermore, the sharpness of our results enables us to present a quantitative and accurate comparison with convex optimization based recovery algorithms [Dhifallah and Lu, 2017; Dhifallah *et al.*, 2017; Abbasi *et al.*, 2017]. We will formally introduce our message passing algorithm and state our main results in Chapter 3.

1.3.1 Background

1.3.1.1 Existing theoretical work

Early theoretical results on phase retrieval, such as PhaseLift [Candès *et al.*, 2013] and PhaseCut [Waldspurger *et al.*, 2015], are based on semidefinite relaxations. For random Gaussian measurements, a variant of PhaseLift can recover the signal exactly (up to global phase) in the noiseless setting using $O(n)$ measurements [Candès and Li, 2014]. However, PhaseLift (or PhaseCut) involves solving a semidefinite programming (SDP) and is computationally prohibitive for large-scale applications. A different convex optimization approach for phase retrieval, which has the same $O(n)$ sample complexity, was independently proposed in Goldstein and Studer [2016] and Bahmani and Romberg [2016]. This method is formulated in the natural signal space and does not involve lifting, and is therefore computationally more attractive than SDP-based counterparts. However, both methods require an anchor vector that has non-zero correlation with the true signal, and the quality of the recovery highly depends on the quality of the anchor.

Apart from convex relaxation approaches, non-convex optimization approaches attract considerable recent interests. These algorithms typically consist of a carefully designed initialization step (usually accomplished via a spectral method [Netrapalli *et al.*, 2013]) followed by iterations that refine the estimate. An early work in this direction is the alternating minimization algorithm proposed in Netrapalli *et al.* [2013], which has sub-optimal sample complexity. Another line of work includes the Wirtinger flow algorithm [Candès *et al.*, 2015; Ma *et al.*, 2017], truncated Wirtinger flow algorithm [Chen and Candès, 2017], and other variants[Cai *et al.*, 2016; Zhang and Liang, 2016; Wang *et al.*, 2016; Soltanolkotabi, 2017]. Other approaches include Kaczmarz method [Wei, 2015; Chi and Lu, 2016; Tan and Vershynin, 2017; Jeong

and Güntürk, 2017], trust region method [Sun *et al.*, 2016], coordinate decent [Zeng and So, 2017], prox-linear algorithm [Duchi and Ruan, 2017] and Polyak subgradient method [Davis *et al.*, 2017].

All the above theoretical results guarantee successful recovery with $m = \delta n$ measurements (or more) where δ is a fixed often large constant. However, such theories are not capable of providing fair comparison among different algorithms. To resolve this issue researchers have started studying the performance of different algorithms under the asymptotic setting $m/n \to \delta$ and $n \to \infty$. An interesting iterative projection method was proposed inLi *et al.* [2015], whose dynamics can be characterized exactly under this asymptotic setting. However, Li *et al.* [2015] does not analyze the number of measurements required for this algorithm to work. The work in Lu and Li [2017] provides sharp characterization of the spectral initialization step (which is a key ingredient to many of the above algorithms). The analysis in Lu and Li [2017] reveals a phase transition phenomenon: spectral method produces an estimate not orthogonal to the signal if and only if δ is larger than a threshold (called "weak threshold" in Mondelli and Montanari [2017]). Later, Mondelli and Montanari [2017] derived the information-theoretically optimal weak threshold (which is 0.5 for the real-valued model and 1 for the complex-valued model) and proved that the optimal weak threshold can be achieved by an optimally-tuned spectral method. Using the non-rigorous replica method from statistical physics, Dhifallah and Lu [2017] analyzes the exact threshold of δ (for the real-value setting) above which the PhaseMax method in Goldstein and Studer [2016] and Bahmani and Romberg [2016] achieves perfect recovery. The analysis in Dhifallah and Lu [2017] shows that the performance of PhaseMax highly depends on initialization (see Fig. 1 of Dhifallah and Lu [2017]), and the required δ is lower bounded by 2 for real-valued models. On the other hand, we will show that AMP.A proposed in this paper achieves perfect recovery for $\delta > 1.5$

under the same setting. The analysis in Dhifallah and Lu [2017] was later rigorously proved in Dhifallah *et al.* [2017] via the Gaussian min-max framework [Thrampoulidis *et al.*, 2016, 2015], and a new algorithm called PhaseLamp was proposed. The PhaseLamp method has superior recovery performance over PhaseMax, but again it does not work when $\delta < 2$ for real-valued models. Further, Dhifallah and Lu [2017]; Dhifallah *et al.* [2017] focus on the noiseless scenario, while in this paper we also analyze the noise sensitivity of AMP.A. Finally, a recent paper Abbasi *et al.* [2017] derived an upper bound of δ such that PhaseLift achieves perfect recovery. The exact value of this upper bound can be derived by solving a three-variable convex optimization problem and empirically Abbasi *et al.* [2017] shows that $\delta \approx 3$ for real-valued models.

1.3.1.2 Existing work based on AMP

Our work in this paper is based on the approximate message passing (AMP) framework [Donoho *et al.*, 2009; Bayati and Montanari, 2011], in particular the generalized approximate message passing (GAMP) algorithm developed and analyzed in Rangan [2011]; Javanmard and Montanari [2013]. A key property of AMP (including GAMP) is that its asymptotic behavior can be characterized exactly via the state evolution platform [Donoho *et al.*, 2009; Bayati and Montanari, 2011; Rangan, 2011; Javanmard and Montanari, 2013].

For phase retrieval, a Bayesian GAMP algorithm has been proposed in Schniter and Rangan [2015]. However, Schniter and Rangan [2015] did not provide rigorous performance analysis, partly due to the heuristic treatments used in the algorithm (such as damping and restart). Another work related to ours is the recent paper Barbier *et al.* [2017], which analyzed the phase transitions of the Bayesian GAMP algorithms for a class of nonlinear acquisition models. For the phase retrieval problem, a phase transition diagram was shown in Barbier *et al.* [2017, Fig. 1] under a

Bernoulli-Gaussian signal prior. The numerical results in Barbier *et al.* [2017] indeed achieve state-of-the-art reconstruction results for real-valued models. However, Barbier *et al.* [2017] did not provide the analysis of their results and in particular did not mention how they handle a difficulty related to initialization. Further, the algorithm in Barbier *et al.* [2017] is based on the Bayesian framework which assumes that the signal and the measurements are generated according to some known distributions. Contrary to Schniter and Rangan [2015] and Barbier *et al.* [2017], this paper considers a version of GAMP derived from solving the popular optimization problem Equation (1.4). We provide rigorous performance analysis of our algorithm for both real and complex-valued models. Note that the advantages and disadvantages of Bayesian and optimization-based techniques have been a long debate in the field of Statistics. Hence, we do not repeat those debates here. Given our experience in the fields of compressed sensing and phase retrieval, it seems that the performance of Bayesian algorithms are more sensitive to their assumptions than the optimization-based schemes. Furthermore, performance analyses of Bayesian algorithms are often very challenging under "non-ideal" situations which the algorithms are not designed for.

Here, we emphasize another advantage of our approach. Given the fact that the most popular schemes in practice are iterative algorithms derived for solving non-convex optimization problems, the detailed analyses of AMP.A presented in our paper may also shed light on the performance of these algorithms and suggest new ideas to improve their performances.

1.3.1.3 Fundamental limits

It the literature of phase retrieval, it is well known that to make the signal-to-observation mapping injective one needs at least $m = 4n$ measurements [Bandeira

et al., 2014] (or $m = 2n$ [Balan *et al.*, 2006] in the case of real-valued models). On the other hand, the measurement thresholds obtained in this paper are $\delta = \frac{64}{\pi^2} - 4 \approx 2.5$ for complex-valued signal and $\delta = \frac{\pi^2}{4} - 1 \approx 1.5$ for real-valued signal respectively. In fact, our algorithm can in principal recover the signal when $\delta > 2$ and $\delta > 1 + \frac{4}{\pi^2}$ (or $\delta > 1$ if continuation is not applied) for complex and real-valued models, provided that the algorithm is initialized close enough to the signal (though no known initialization strategy can accomplish this goal). Hence, our threshold are even smaller than the injectivity bounds. We emphasize that this is possible since the injectivity bounds derived in Balan *et al.* [2006]; Bandeira *et al.* [2014] are defined for *all* $\boldsymbol{x}^\star$ (which can depend on $\boldsymbol{A}$ in the worst case scenario). This is different from our assumption that $\boldsymbol{x}^\star$ is independent of $\boldsymbol{A}$, which is more relevant in applications where one has some freedom to randomize the sampling mechanism. In fact, several papers have observed that their algorithm can operate at the injectivity thresholds $\delta = 2$ for real-valued models [Wang *et al.*, 2016; Duchi and Ruan, 2017]. These two different notions of thresholds were discussed in Jalali and Maleki [2016]. In the context of phase retrieval, the reader is referred to the recent paper Bakhshizadeh *et al.* [2017], which showed that by solving a compression-based optimization problem, the required number of observations for recovery is essentially the information dimension of the signal (see Bakhshizadeh *et al.* [2017] for the precise definition). For instance, if the signal is k-sparse and complex-valued, then $2k$ measurements suffice.

1.4 Dynamical System

A dynamical system is a system in which a function describes the time dependence of a point in a geometrical space. In general, the time dependence can be defined on real line, yet in this thesis, we focus on the case when it is restricted on non-nagative

integers. Formally, we focus on the following dynamical system $\{\boldsymbol{\eta}^{\langle t \rangle} \in \mathbb{R}^d\}_{t \geq 0}$ defined by a continuous update rule $f : \mathbb{R}^d \mapsto \mathbb{R}^d$, i.e.,

$$\boldsymbol{\eta}^{\langle t+1 \rangle} = f(\boldsymbol{\eta}^{\langle t \rangle}).$$

By definition, many iterative algorithms are also dynamical systems, and therefore studying the convergence of the algorithms is equivalent to study the trajectories of $\boldsymbol{\eta}^{\langle t \rangle}$ for the dynamical systems. It is easy to show that when $\boldsymbol{\eta}^{\langle t \rangle}$ converges, the converging point should be one of the solutions to the equation $f(\boldsymbol{\eta}) = \boldsymbol{\eta}$ which are referred as *the fixed points* of this dynamical system. However, except for some special cases such as linear dynamical systems, it is hard in general to show whether $\boldsymbol{\eta}^{\langle t \rangle}$ converges, and if so which fixed point $\boldsymbol{\eta}^{\langle t \rangle}$ converges to when there are more than one fixed point. A standard approach to prove the convergence of $\boldsymbol{\eta}^{\langle t \rangle}$ to a fixed point $\boldsymbol{\eta}^\star$ is to establish the following inequality for some constants $\rho \in (0, 1)$:

$$\|f(\boldsymbol{\eta}) - \boldsymbol{\eta}^\star\|_2^2 \leq \rho \|\boldsymbol{\eta} - \boldsymbol{\eta}^\star\|_2^2 . \tag{1.5}$$

In Section 2.4, we will prove Equation (1.5) for a class of one dimensional dynamical systems. However for many cases, Equation (1.5) only holds in a small neighborhood around $\boldsymbol{\eta}^\star$ and the trajectories become unclear for initializations outside the neighborhood. A general approach is to replace the ℓ_2 distance in Equation (1.5) with any continuous non-negative function M with the property that $\boldsymbol{\eta}^\star$ is the unique solution to $M(\boldsymbol{\eta}) = 0$. Yet, finding such function M remains a big challenge. In this thesis, we propose a geometric approach to construct the function M for a class of two dimensional dynamical systems, and prove convergence of $\boldsymbol{\eta}^{\langle t \rangle}$ to the desired fixed point $\boldsymbol{\eta}^\star$.

Before we proceed to the main results and their proofs in the next chapters, let us specify the following important concept of stable fixed points.

Definition 1. Let $\eta^\star$ be a fixed point of an one-dimensional dynamical system defined by the update rule $f : \mathbb{R} \mapsto \mathbb{R}$. We say $\eta^\star$ is a *stable fixed point* if there exists $\epsilon > 0$ such that either

$$\eta^{\langle t \rangle} \to \eta^\star, \qquad \forall \eta^{\langle 0 \rangle} \in (\eta^\star - \epsilon, \eta^\star)$$

or

$$\eta^{\langle t \rangle} \to \eta^\star, \qquad \forall \eta^{\langle 0 \rangle} \in (\eta^\star, \eta^\star + \epsilon)$$

holds.

In other words, a fixed point is considered not stable if a small perturbation will let the trajectory avoid the fixed point. There are other notions that discuss the stability of a fixed point and this concept can be extended to multi-dimension and. We focus on our definition in one dimension because of two reasons. First, in many cases, verifying a stable fixed point is relatively easy in one dimension. The following cases are such examples,

(i) There exists $\epsilon > 0$ such that $\eta^\star \geq f(\eta) > \eta$ for all $\eta \in (\eta^\star - \epsilon, \eta^\star)$ or $\eta^\star \leq f(\eta) < \eta$ for all $\eta \in (\eta^\star, \eta^\star + \epsilon)$. Then the fixed point $\eta^\star$ is stable.

(ii) There exists $\epsilon > 0$ such that $f(\eta) < \eta$ for all $\eta \in (\eta^\star - \epsilon, \eta^\star)$ and $f(\eta) > \eta$ for all $\eta \in (\eta^\star, \eta^\star + \epsilon)$. Then the fixed point $\eta^\star$ is not stable.

(iii) The function f is Lipchitz in a neighborhood around $\eta^\star$ with Lipchitz constant strictly smaller than 1.

The proof for the last two examples are straightforward since they directly imply whether Equation (1.5) holds or not. To prove the first example, note that $\{\eta^{\langle t \rangle}\}$ is a

monotonic sequence in either $(\eta^\star - \epsilon, \eta^\star)$ or $(\eta^\star, \eta^\star + \epsilon)$, and therefore $\eta^{\langle t \rangle}$ converges. Since $\eta^\star$ is the only fixed point in $(\eta^\star - \epsilon, \eta^\star]$ or $[\eta^\star, \eta^\star + \epsilon)$, $\eta^{\langle t \rangle}$ should converge to $\eta^\star$.

Secondly, we can exploit the relation between stable fixed points and the inequality $f(\eta) \gtrless \eta$ in one dimension. Example (i) above shows when the inequality can imply stable fixed points. On the other hand, suppose $\eta_l < \eta_r$ are the only two fixed points in $[\eta_l, \eta_r]$ and f is monotonic on $[\eta_l, \eta_r]$, then we have

- If η_l is not a stable fixed point, then $f(\eta) > \eta$ for all $\eta \in (\eta_l, \eta_r)$.
- If η_r is not a stable fixed point, then $f(\eta) < \eta$ for all $\eta \in (\eta_l, \eta_r)$.

The above claims can be proved by simple contradiction arguments. Note that inequality $f(\eta) \gtrless \eta$ implies the direction of the movement from $\eta^{\langle t \rangle}$ to $\eta^{\langle t+1 \rangle}$. When $\boldsymbol{\eta}^{\langle t \rangle}$ is multi-dimensional vector, we combine the directions of the movement in all axes and deduct the trajectory of $\boldsymbol{\eta}^{\langle t \rangle}$. See our proofs for more details.

1.5 Notations

Since maximum likelihood estimation for GMM and phase retrieval are two separate non-convex optimization problems, we will use two separate sets of notations for each problem with details in Chapter 2 and 3 respectively. Here we only summarize the common notations for complex number, vectors and matrices.

$\bar{a}$ denotes the conjugate of a complex number a. $\angle a$ denotes the phase of a. We use bold lower-case and upper case letters for vectors and matrices respectively. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{H}}$ denote the transpose of a matrix and its Hermitian respectively. Throughout the thesis, we also use the following two notations: $\mathbf{1} \triangleq [1, \ldots, 1]^{\mathrm{T}}$ and $\mathbf{0} \triangleq [0, \ldots, 0]^{\mathrm{T}}$. $\phi(x)$ and $\Phi(x)$ are used for the probability density function and

cumulative distribution function of the standard Gaussian random variable. A random variable a said to be circularly-symmetric Gaussian, denoted as $a \sim \mathcal{CN}(0, \sigma^2)$, if $a = a_R + \mathrm{i}a_I$ and a_R and a_I are two independent real Gaussian random variables with mean zero and variance $\sigma^2/2$. Finally, we define $\langle \boldsymbol{a}, \boldsymbol{b} \rangle \triangleq \sum_{i=1} \bar{a}_i b_i$ for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^d$ or $\langle \boldsymbol{a}, \boldsymbol{b} \rangle \triangleq \sum_{i=1} a_i b_i$ for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$.

Chapter 2

Expectation Maximization for Gaussian Mixture Models

In this chapter, we study the Expectation Maximization algorithm in the context of some simple yet popular and well-studied Gaussian mixture models specifically mixtures of two Gaussians. We want to address the questions about the performance such as the convergence rate and accuracy of the algorithm. Towards this goal, we study an idealized execution of EM in the large sample limit, where the E-step is modified to be computed over an infinitely large i.i.d. sample from a Gaussian mixture distribution in the model. In effect, in the formula for $\hat{Q}(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{\langle t \rangle})$, we replace the observed data $\mathcal{Y}$ with a random variable $\boldsymbol{Y} \sim f(\boldsymbol{y}; \boldsymbol{\theta}^\star)$ for some Gaussian mixture parameters $\boldsymbol{\theta}^\star$ and then take its expectation. The resulting E- and M-steps in iteration t are

$$\text{E-step:} \qquad Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{\langle t \rangle}) \triangleq \mathbb{E}_{\boldsymbol{Y}} \left[\sum_{\boldsymbol{z}} f(\boldsymbol{z} \mid \boldsymbol{Y}; \boldsymbol{\theta}^{\langle t \rangle}) \log f(\boldsymbol{Y}, \boldsymbol{z}; \boldsymbol{\theta}) \right], \tag{2.1}$$

$$\text{M-step:} \qquad \boldsymbol{\theta}^{\langle t+1 \rangle} \triangleq \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{\langle t \rangle}). \tag{2.2}$$

This sequence of parameters $(\boldsymbol{\theta}^{\langle t \rangle})_{t \geq 0}$ is fully determined by the initial setting $\boldsymbol{\theta}^{\langle 0 \rangle}$. We refer to this idealization as *Population EM*. (To avoid confusion, we refer the original EM algorithm run with a finite sample as *Sample-based EM*.) Not only does Population EM shed light on the dynamics of EM in the large sample limit, but it can also reveal some of the fundamental limitations of EM. Indeed, if Population EM cannot provide an accurate estimate for the parameters $\boldsymbol{\theta}^\star$, then intuitively, one would not expect the EM algorithm with a finite sample size to do so either.

The structure of this chapter is the following: we first introduce our models in Section 2.1. Then, we present results based on Population EM and Sample-based EM in Section 2.2 and Section 2.3 respectively. Next, we present proofs for these results in Section 2.4 and Section 2.5 respectively. Finally we present some interesting auxiliary results in Section 2.6.

Notation: Let $\phi_d(\mu, \boldsymbol{\Sigma})$ be the pdf for the general Gaussian distribution with mean μ and covariance $\boldsymbol{\Sigma}$. We use $\phi(x)$ be the pdf of the standard normal distribution $\mathcal{N}(0, 1)$. We use $\Phi(x)$ be the CDF of the standard normal distribution $\mathcal{N}(0, 1)$. Throughout, we denote the Euclidean norm by $\| \cdot \|$, and the signum function by $\operatorname{sgn}(\cdot)$ (where $\operatorname{sgn}(0) = 0$, $\operatorname{sgn}(z) = 1$ if $z > 0$, and $\operatorname{sgn}(z) = -1$ if $z < 0$). Finally, we use Population $\text{EM}_{\#}$ as a short hand for the Population EM iterates with Model $\#$.

2.1 Models

In this section, we introduce the specific models we study in this chapter, along with the corresponding Sample-based EM and Population EM updates.

Model 1. The observation $\mathcal{Y}$ is an i.i.d. sample from the mixture distribution $0.5\mathcal{N}(-\boldsymbol{\theta}^\star, \boldsymbol{\Sigma}) + 0.5\mathcal{N}(\boldsymbol{\theta}^\star, \boldsymbol{\Sigma})$; $\boldsymbol{\Sigma}$ is a known covariance matrix in $\mathbb{R}^d$, and $\boldsymbol{\theta}^\star$ is the unknown parameter of interest.

1. Sample-based EM iteratively updates its estimate of $\boldsymbol{\theta}^\star$ according to the following equation:

$$\hat{\boldsymbol{\theta}}^{\langle t+1\rangle} \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\Big(2\mathsf{w}_d(\boldsymbol{y}_i, \hat{\boldsymbol{\theta}}^{\langle t\rangle}) - 1\Big)\,\boldsymbol{y}_i, \tag{2.3}$$

   where $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are the independent draws that comprise $\mathcal{Y}$,

$$\mathsf{w}_d(\boldsymbol{y}, \boldsymbol{\theta}) \quad \triangleq \quad \frac{\phi_d(\boldsymbol{y} - \boldsymbol{\theta}, \boldsymbol{\Sigma})}{\phi_d(\boldsymbol{y} - \boldsymbol{\theta}, \boldsymbol{\Sigma}) + \phi_d(\boldsymbol{y} + \boldsymbol{\theta}, \boldsymbol{\Sigma})}.$$

2. Population EM iteratively updates its estimate according to the following equation:

$$\boldsymbol{\theta}^{\langle t+1\rangle} \quad = \quad \mathbb{E}_{\boldsymbol{y}\sim f_1^\star}(2\mathsf{w}_d(\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle}) - 1)\boldsymbol{y}, \tag{2.4}$$

   where $f_1^\star = f_1^\star(\boldsymbol{\theta}^\star)$ here denotes the true distribution $\frac{1}{2}\mathcal{N}(-\boldsymbol{\theta}^\star, \boldsymbol{\Sigma}) + \frac{1}{2}\mathcal{N}(\boldsymbol{\theta}^\star, \boldsymbol{\Sigma})$.

Our first model is also studied in Balakrishnan *et al.* [2017]; Daskalakis *et al.* [2017]. Note that, we impose equal weights, symmetric means conditions on Model 1. To generalize Model 1, we relax one of these conditions separately and obtain the following 3 models.

Model 2. The observation $\mathcal{Y}$ is an i.i.d. sample from the mixture distribution $0.5\mathcal{N}(\boldsymbol{\mu}_1^\star, \boldsymbol{\Sigma}) + 0.5\mathcal{N}(\boldsymbol{\mu}_2^\star, \boldsymbol{\Sigma})$. Again, $\boldsymbol{\Sigma}$ is known, and $(\boldsymbol{\mu}_1^\star, \boldsymbol{\mu}_2^\star)$ are the unknown parameters of interest.

1. Sample-based EM iteratively updates its estimate of $\boldsymbol{\mu}_1^\star$ and $\boldsymbol{\mu}_2^\star$ at every iteration according to the following equations:

$$\hat{\boldsymbol{\mu}}_1^{\langle t+1\rangle} = \frac{\sum_{i=1}^n \mathsf{v}_d(\boldsymbol{y}_i, \hat{\boldsymbol{\mu}}_1^{\langle t\rangle}, \hat{\boldsymbol{\mu}}_2^{\langle t\rangle})\boldsymbol{y}_i}{\sum_{i=1}^n \mathsf{v}_d(\boldsymbol{y}_i, \hat{\boldsymbol{\mu}}_1^{\langle t\rangle}, \hat{\boldsymbol{\mu}}_2^{\langle t\rangle})}, \tag{2.5}$$

$$\hat{\boldsymbol{\mu}}_2^{\langle t+1\rangle} = \frac{\sum_{i=1}^n (1 - \mathsf{v}_d(\boldsymbol{y}_i, \hat{\boldsymbol{\mu}}_1^{\langle t\rangle}, \hat{\boldsymbol{\mu}}_2^{\langle t\rangle}))\boldsymbol{y}_i}{\sum_{i=1}^n (1 - \mathsf{v}_d(\boldsymbol{y}_i, \hat{\boldsymbol{\mu}}_1^{\langle t\rangle}, \hat{\boldsymbol{\mu}}_2^{\langle t\rangle}))}, \tag{2.6}$$

   where $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are the independent draws that comprise $\mathcal{Y}$, and

$$\mathsf{v}_d(\boldsymbol{y}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \triangleq \frac{\phi_d(\boldsymbol{y} - \boldsymbol{\mu}_1, \boldsymbol{\Sigma})}{\phi_d(\boldsymbol{y} - \boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + \phi_d(\boldsymbol{y} - \boldsymbol{\mu}_2, \boldsymbol{\Sigma})}.$$

2. Population EM iteratively updates its estimates according to the following equations:

$$\boldsymbol{\mu}_1^{\langle t+1\rangle} = \frac{\mathbb{E}_{\boldsymbol{y}\sim f_2^\star} \mathsf{v}_d(\boldsymbol{y}, \boldsymbol{\mu}_1^{\langle t\rangle}, \boldsymbol{\mu}_2^{\langle t\rangle})\boldsymbol{y}}{\mathbb{E}\mathsf{v}_d(\boldsymbol{y}, \boldsymbol{\mu}_1^{\langle t\rangle}, \boldsymbol{\mu}_2^{\langle t\rangle})}, \tag{2.7}$$

$$\boldsymbol{\mu}_2^{\langle t+1\rangle} = \frac{\mathbb{E}_{\boldsymbol{y}\sim f_2^\star} (1 - \mathsf{v}_d(\boldsymbol{y}, \boldsymbol{\mu}_1^{\langle t\rangle}, \boldsymbol{\mu}_2^{\langle t\rangle}))\boldsymbol{y}}{\mathbb{E}(1 - \mathsf{v}_d(\boldsymbol{y}, \boldsymbol{\mu}_1^{\langle t\rangle}, \boldsymbol{\mu}_2^{\langle t\rangle}))}, \tag{2.8}$$

   where $f_2^\star = f_2^\star(\boldsymbol{\mu}_1^\star, \boldsymbol{\mu}_2^\star)$ here denotes the true distribution $\frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_1^\star, \boldsymbol{\Sigma}) + \frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_2^\star, \boldsymbol{\Sigma})$.

Model 3. The observation $\mathcal{Y}$ is an i.i.d. sample from the mixture distribution $w_1^\star\mathcal{N}(\boldsymbol{\theta}^\star, \boldsymbol{\Sigma}) + w_2^\star\mathcal{N}(-\boldsymbol{\theta}^\star, \boldsymbol{\Sigma})$. Both covariance matrix $\boldsymbol{\Sigma}$ and weights $w_1^\star, w_2^\star$ are known. The mean $\boldsymbol{\theta}^\star$ is the unknown parameter of interest.

1. Sample-based EM iteratively updates its estimate of $\boldsymbol{\theta}^\star$ according to the follow-

ing equation:

$$\hat{\boldsymbol{\theta}}^{\langle t+1\rangle} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{w_1^\star\phi_d(\boldsymbol{y}_i-\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})-w_2^\star\phi_d(\boldsymbol{y}_i+\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})}{w_1^\star\phi_d(\boldsymbol{y}_i-\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})+w_2^\star\phi_d(\boldsymbol{y}_i+\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})}\boldsymbol{y}_i\right]. \quad (2.9)$$

2. Population EM iteratively updates its estimates according to the following equations:

$$\boldsymbol{\theta}^{\langle t+1\rangle} = \mathbb{E}_{\boldsymbol{y}\sim f_3^\star}\left[\frac{w_1^\star\phi_d(\boldsymbol{y}-\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})-w_2^\star\phi_d(\boldsymbol{y}+\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})}{w_1^\star\phi_d(\boldsymbol{y}-\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})+w_2^\star\phi_d(\boldsymbol{y}+\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})}\boldsymbol{y}\right], \quad (2.10)$$

where $f_3^\star = f_3^\star(\boldsymbol{\theta}^\star, w_1^\star)$ here denotes the true distribution

$$w_1^\star\mathcal{N}(\boldsymbol{\theta}^\star,\boldsymbol{\Sigma}) + w_2^\star\mathcal{N}(-\boldsymbol{\theta}^\star,\boldsymbol{\Sigma}).$$

Model 4. The observation $\mathcal{Y}$ is an i.i.d. sample from the mixture distribution $w_1^\star\mathcal{N}(\boldsymbol{\theta}^\star,\boldsymbol{\Sigma}) + w_2^\star\mathcal{N}(-\boldsymbol{\theta}^\star,\boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$ is known. Both the weights $w_1^\star, w_2^\star$ and the mean $\boldsymbol{\theta}^\star$ are the unknown parameters of interest.

1. Sample-based EM iteratively updates its estimates of $w_1^\star, w_2^\star, \boldsymbol{\theta}^\star$ according to the following equation:

$$\begin{aligned}\hat{w}_1^{\langle t+1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{w}_1^{\langle t\rangle}\phi_d(\boldsymbol{y}_i-\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})}{\hat{w}_1^{\langle t\rangle}\phi_d(\boldsymbol{y}_i-\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})+\hat{w}_2^{\langle t\rangle}\phi_d(\boldsymbol{y}_i+\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})}\right] = 1-\hat{w}_2^{\langle t+1\rangle}.\\ \hat{\boldsymbol{\theta}}^{\langle t+1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{w}_1^{\langle t\rangle}\phi_d(\boldsymbol{y}_i-\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})-\hat{w}_2^{\langle t\rangle}\phi_d(\boldsymbol{y}_i+\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})}{\hat{w}_1^{\langle t\rangle}\phi_d(\boldsymbol{y}_i-\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})+\hat{w}_2^{\langle t\rangle}\phi_d(\boldsymbol{y}_i+\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})}\boldsymbol{y}_i\right]. \quad (2.11)\end{aligned}$$

2. Population EM iteratively updates its estimates according to the following equa-

tions:

$$
\begin{aligned}
w_1^{\langle t+1\rangle} &= \mathbb{E}_{\boldsymbol{y}\sim f_4^\star}\left[\frac{w_1^{\langle t\rangle}\phi_d(\boldsymbol{y}-\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})}{w_1^{\langle t\rangle}\phi_d(\boldsymbol{y}-\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})+w_2^{\langle t\rangle}\phi_d(\boldsymbol{y}+\hat{\boldsymbol{\theta}}^{\langle t\rangle},\boldsymbol{\Sigma})}\right] \qquad (2.12)\\
&= 1-w_2^{\langle t+1\rangle}.\\
\boldsymbol{\theta}^{\langle t+1\rangle} &= \mathbb{E}_{\boldsymbol{y}\sim f_4^\star}\left[\frac{w_1^{\langle t\rangle}\phi_d(\boldsymbol{y}-\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})-w_2^{\langle t\rangle}\phi_d(\boldsymbol{y}+\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})}{w_1^{\langle t\rangle}\phi_d(\boldsymbol{y}-\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})+w_2^{\langle t\rangle}\phi_d(\boldsymbol{y}+\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\Sigma})}\boldsymbol{y}\right]. \qquad (2.13)
\end{aligned}
$$

where $f_4^\star = f_4^\star(\boldsymbol{\theta}^\star, w_1^\star)$ here denotes the true distribution

$$w_1^\star\mathcal{N}(\boldsymbol{\theta}^\star,\boldsymbol{\Sigma})+w_2^\star\mathcal{N}(-\boldsymbol{\theta}^\star,\boldsymbol{\Sigma}).$$

One can check that Model 2 to 4 are extensions of Model 1. For example, to verify for Model 2, let us consider the following re-parametrization of the model parameters and Population EM iterates:

$$\boldsymbol{a}^{\langle t\rangle} \triangleq \frac{\boldsymbol{\mu}_1^{\langle t\rangle}+\boldsymbol{\mu}_2^{\langle t\rangle}}{2}-\frac{\boldsymbol{\mu}_1^\star+\boldsymbol{\mu}_2^\star}{2}, \quad \boldsymbol{\theta}^{\langle t\rangle} \triangleq \frac{\boldsymbol{\mu}_2^{\langle t\rangle}-\boldsymbol{\mu}_1^{\langle t\rangle}}{2}, \quad \boldsymbol{\theta}^\star \triangleq \frac{\boldsymbol{\mu}_2^\star-\boldsymbol{\mu}_1^\star}{2}. \qquad (2.14)$$

We will explain why we abuse the notation of $\boldsymbol{\theta}^{\langle t\rangle}$ and $\boldsymbol{\theta}^\star$ shortly. The iterations of Population EM can be written in terms of these new parameters $\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}$ as

$$
\begin{aligned}
\boldsymbol{a}^{\langle t+1\rangle} &= \frac{\boldsymbol{\gamma}^{\langle t+1\rangle}(1-2\mathsf{p}^{\langle t+1\rangle})}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})}, \qquad (2.15)\\
\boldsymbol{\theta}^{\langle t+1\rangle} &= \frac{\boldsymbol{\gamma}^{\langle t+1\rangle}}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})}, \qquad (2.16)
\end{aligned}
$$

where

$$\begin{aligned} \boldsymbol{\gamma}^{\langle t+1\rangle} &= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star} \mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} \\ &, \\ \mathsf{p}^{\langle t+1\rangle} &= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star} \mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\,. \end{aligned} \tag{2.17}$$

The following lemma establishes a connection between the iterations of Population EM for Model 1 and for Model 2.

Lemma 2.1. Suppose $f_1^\star$ and $f_2^\star$ correspond to the same distribution, i.e, $\boldsymbol{\mu}_1^\star = -\boldsymbol{\mu}_2^\star = \boldsymbol{\theta}^\star$. Then for Population EM_2, if $\boldsymbol{a}^{\langle 0\rangle} = \mathbf{0}$, then $\boldsymbol{a}^{\langle t\rangle} = \mathbf{0}$ for every t. Furthermore,

$$\boldsymbol{\theta}^{\langle t+1\rangle} \;=\; 2\mathbb{E}_{\boldsymbol{y}\sim f_2^\star} \mathsf{w}_d(\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} \;=\; \mathbb{E}_{\boldsymbol{y}\sim f_1^\star}(2\mathsf{w}_d(\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle}) - 1)\boldsymbol{y}\,.$$

Proof. The proof is a simple induction that exploits the fact that

$$\mathsf{w}_d(\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle}) + \mathsf{w}_d(-\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle}) = 1.$$

That is if $\boldsymbol{a}^{\langle t\rangle} = \mathbf{0}$, then

$$\mathsf{p}^{\langle t+1\rangle} \;=\; \mathbb{E}_{\boldsymbol{y}\sim f_2^\star} \mathsf{w}_d(\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle}) \;=\; \frac{1}{2}(\mathbb{E}_{\boldsymbol{y}\sim f_2^\star} \mathsf{w}_d(\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle}) + \mathbb{E}_{\boldsymbol{y}\sim f_2^\star} \mathsf{w}_d(-\boldsymbol{y}, \boldsymbol{\theta}^{\langle t\rangle})) \;=\; \frac{1}{2}\,.$$

□

Observe that the expression for $\boldsymbol{\theta}^{\langle t+1\rangle}$ in Lemma 2.1 is the same as the Population EM update under Model 1, given in Equation (2.4). Therefore, Lemma 2.1 tells us that Model 1 is a special case of Model 2 if we know the value of the mean $(\boldsymbol{\mu}_1^\star + \boldsymbol{\mu}_2^\star)/2$. In this case, $\boldsymbol{\theta}^{\langle t\rangle}$ is regarded as an estimate of $(\boldsymbol{\mu}_2^\star - \boldsymbol{\mu}_1^\star)/2$, in the same

way that $\boldsymbol{\theta}^{\langle t \rangle}$ is an estimate of $\boldsymbol{\theta}^\star$ in Model 1. This explains our choice of the notation $\boldsymbol{\theta}^{\langle t \rangle} \triangleq (\boldsymbol{\mu}_2^{\langle t \rangle} - \boldsymbol{\mu}_1^{\langle t \rangle})/2$ and $\boldsymbol{\theta}^\star \triangleq (\boldsymbol{\mu}_2^\star - \boldsymbol{\mu}_1^\star)/2$ in Equation (2.14).

2.2 Main Results for Population EM

In this section, we present our main results for Population EM with all models. These results are based on the two paper Xu *et al.* [2016] and Xu *et al.* [2018]. We present the proofs in Section 2.4 with more details left in Appendix A.1.

Note that due to the property of Gaussian distribution and the EM algorithm for GMM (See Lemma 2.3), we may assume that the known covariance matrix $\boldsymbol{\Sigma}$ is the identity matrix $\boldsymbol{I}_d$ without loss of generality. Hence, for simplification, we present our main results for the case when $\boldsymbol{\Sigma} = \boldsymbol{I}_d$ for all models. We start with Model 1.

Theorem 2.1. Assume $\boldsymbol{\theta}^\star \in \mathbb{R}^d \setminus \{\mathbf{0}\}$. Let $(\boldsymbol{\theta}^{\langle t \rangle})_{t \geq 0}$ denote the Population EM iterates for Model 1.

- If $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle = 0$, then

$$\boldsymbol{\theta}^{\langle t \rangle} \;\; \to \;\; \mathbf{0} \quad \text{as } t \to \infty \,.$$

- If $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle \neq 0$, then there exists $\kappa_\theta \in (0, 1)$—depending only on $\boldsymbol{\theta}^\star$ and $\boldsymbol{\theta}^{\langle 0 \rangle}$—such that

$$\left\| \boldsymbol{\theta}^{\langle t+1 \rangle} - \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle) \boldsymbol{\theta}^\star \right\| \;\; \leq \;\; \kappa_\theta \cdot \left\| \boldsymbol{\theta}^{\langle t \rangle} - \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle) \boldsymbol{\theta}^\star \right\| .$$

Theorem 2.1 asserts that the sequence $(\boldsymbol{\theta}^{\langle t \rangle})_{t \geq 0}$ converges to $\operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle) \boldsymbol{\theta}^\star$. Further, if $\boldsymbol{\theta}^{\langle 0 \rangle}$ is not on the hyperplane $\{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{\theta}^\star \rangle = 0\}$, the EM algorithm

finds either $\boldsymbol{\theta}^\star$ or $-\boldsymbol{\theta}^\star$, which both are the global optimum, at a linear convergence rate.

We now discuss Population EM with Model 2 which is an extension of Model 1 when the symmetric condition on the mean parameter is relaxed. To state our results more concisely, we use the re-parameterization introduced in Equation (2.14). If the sequence of Population EM_2 iterates $((\boldsymbol{\mu}_1^{\langle t \rangle}, \boldsymbol{\mu}_2^{\langle t \rangle}))_{t \geq 0}$ converges to $(\boldsymbol{\mu}_1^\star, \boldsymbol{\mu}_2^\star)$, then we expect $\boldsymbol{\theta}^{\langle t \rangle} \to \boldsymbol{\theta}^\star$. Hence, we also define $\beta^{\langle t \rangle}$ as the angle between $\boldsymbol{\theta}^{\langle t \rangle}$ and $\boldsymbol{\theta}^\star$, i.e.,

$$\beta^{\langle t \rangle} \;\triangleq\; \arccos\left(\frac{\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^\star \rangle}{\|\boldsymbol{\theta}^{\langle t \rangle}\| \|\boldsymbol{\theta}^\star\|}\right) \;\in\; [0, \pi]\,. \tag{2.18}$$

(This is well-defined as long as $\boldsymbol{\theta}^{\langle t \rangle} \neq \mathbf{0}$ and $\boldsymbol{\theta}^\star \neq \mathbf{0}$.)

We now present results on Population EM with Model 2.

Theorem 2.2. Assume $\boldsymbol{\theta}^\star \in \mathbb{R}^d \setminus \{\mathbf{0}\}$. Let $(\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle})_{t \geq 0}$ denote the (re-parameterized) Population EM iterates for Model 2.

- If $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle = 0$, then

$$(\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle}) \;\to\; (\mathbf{0}, \mathbf{0}) \quad \text{as } t \to \infty\,.$$

- If $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle \neq 0$, then $\boldsymbol{\theta}^{\langle t \rangle} \neq \mathbf{0}$ for all $t \geq 0$. Furthermore, there exist $\kappa_a \in (0,1)$ and $\kappa_\beta \in (0,1)$—all depending only on $\boldsymbol{\theta}^\star$ and initialization $(\boldsymbol{a}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\langle 0 \rangle})$—such that

$$\|\boldsymbol{a}^{\langle t+1 \rangle}\|^2 \;\leq\; \kappa_a^2 \cdot \|\boldsymbol{a}^{\langle t \rangle}\|^2 + \frac{\|\boldsymbol{\theta}^\star\|^2 \sin^2(\beta^{\langle t \rangle})}{4}\,, \tag{2.19}$$

$$\sin(\beta^{\langle t+1 \rangle}) \;\leq\; \kappa_\beta^t \cdot \sin(\beta^{\langle 0 \rangle})\,. \tag{2.20}$$

  Finally, there exist $T_0 > 0$, $\kappa_\theta \in (0,1)$, and $c_\theta > 0$—all depending only on $\boldsymbol{\theta}^\star$

and initialization $(\boldsymbol{a}^{\langle 0\rangle}, \boldsymbol{\theta}^{\langle 0\rangle})$—such that

$$\begin{aligned}\left\|\boldsymbol{\theta}^{\langle t+1\rangle} - \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\theta}^{\star}\rangle)\boldsymbol{\theta}^{\star}\right\|^2 \;\leq\; & c_\theta \cdot \|\boldsymbol{a}^{\langle t\rangle}\| \\ & + \kappa_\theta^2 \cdot \left\|\boldsymbol{\theta}^{\langle t\rangle} - \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\theta}^{\star}\rangle)\boldsymbol{\theta}^{\star}\right\|^2 \quad \forall t > T_0\,.\end{aligned} \tag{2.21}$$

Theorem 2.2 implies that

$$\begin{aligned}\boldsymbol{a}^{\langle t\rangle} &\to \mathbf{0} \quad \text{as } t \to \infty\,, \\ \boldsymbol{\theta}^{\langle t\rangle} &\to \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\mu}_2^{\star} - \boldsymbol{\mu}_1^{\star}\rangle)\frac{\boldsymbol{\mu}_2^{\star} - \boldsymbol{\mu}_1^{\star}}{2} \quad \text{as } t \to \infty\,.\end{aligned}$$

Further, when $\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\theta}^{\star}\rangle \neq 0$, by combining the inequalities from Theorem 2.2, we conclude

$$\begin{aligned}\|\boldsymbol{a}^{\langle t+1\rangle}\|^2 &= \kappa_a^{2t}\|\boldsymbol{a}^{\langle 0\rangle}\|^2 + \frac{\|\boldsymbol{\theta}^{\star}\|^2}{4}\sum_{\tau=0}^{t}\kappa_a^{2\tau}\cdot\sin^2(\beta^{\langle t-\tau\rangle}) \\ &\leq \kappa_a^{2t}\|\boldsymbol{a}^{\langle 0\rangle}\|^2 + \frac{\|\boldsymbol{\theta}^{\star}\|^2}{4}\sum_{\tau=0}^{t}\kappa_a^{2\tau}\kappa_\beta^{2(t-\tau)}\cdot\sin^2(\beta^{\langle 0\rangle}) \\ &\leq \kappa_a^{2t}\|\boldsymbol{a}^{\langle 0\rangle}\|^2 + \frac{\|\boldsymbol{\theta}^{\star}\|^2}{4}t\left(\max(\kappa_a,\kappa_\beta)\right)^t\sin^2(\beta^{\langle 0\rangle})\,.\end{aligned}$$

Hence, $\boldsymbol{a}^{\langle t\rangle}$ converges to $\mathbf{0}$ at a linear rate. Similarly, we have $\boldsymbol{\theta}^{\langle t\rangle}$ converges to $\operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\mu}_2^{\star} - \boldsymbol{\mu}_1^{\star}\rangle)\frac{\boldsymbol{\mu}_2^{\star} - \boldsymbol{\mu}_1^{\star}}{2}$ at a linear rate as well. Therefore, Theorem 2.2 shows that the re-parameterized Population EM iterates converge, at a linear rate, to the global optimum as long as the initialization $\boldsymbol{\theta}^{\langle 0\rangle}$ is not orthogonal to $\boldsymbol{\theta}^{\star}$.

We now move to Model 3 and Model 4 which are extensions of Model 1 when the condition of equal weights is relaxed. We first discuss Population EM with Model 3 and show that the same global convergence for Model 1 and 2 may *not* hold for

Model 3 when $w_1^\star \neq w_2^\star$.

Theorem 2.3. Consider Model 3 in dimension one (i.e., $\theta^\star \in \mathbb{R}$). For any $\theta^\star > 0$, there exists $\delta > 0$, such that given $w_1^\star \in (0.5, 0.5 + \delta)$ and initialization $\boldsymbol{\theta}^{\langle 0 \rangle} \leq -\theta^\star$, the Population EM estimate $\boldsymbol{\theta}^{\langle t \rangle}$ for Model 3 converges to a fixed point θ_{wrong} inside $(-\theta^\star, 0)$.

Theorem 2.3 implies that if we use random initialization, the iterates of Population EM with Model 3 may converge to the wrong fixed point with constant probability. We illustrate this in Figure 2.1. We define function $\theta \mapsto H(\theta; \theta^\star, w_1^\star)$ be the update function defined in Equation (2.10), i.e., $\theta^{\langle t+1 \rangle} = H(\theta^{\langle t \rangle}; \theta^\star, w_1^\star)$. Then, the iterates of Population EM with Model 3 converge to a fixed point of H. We have plotted this function for several different values of $w_1^\star$ in the left panel of Figure 2.1. When $w_1^\star$ is close to 1, $H(\theta; \theta^\star, w_1^\star)$ has only one fixed point and that is at $\theta = \theta^\star$. Hence, in this case, the estimates produced by Population EM with Model 3 converge to the true $\theta^\star$. However, when we decrease the value of $w_1^\star$ below a certain threshold (which is numerically found to be approximately 0.77 for $\theta^\star = 1$), two other fixed points of $H(\theta; \theta^\star, w_1^\star)$ emerge. These new fixed points are foils for Population EM with Model 3.

From the failure of Population EM with Model 3, one may expect the over-parameterized Model 4 to fail as well. Yet, surprisingly, our next theorem proves the opposite is true: Population EM with Model 4 has global convergence even when $w_1^\star \neq w_2^\star$.

Theorem 2.4. For any $w_1^\star \in [0.5, 1)$, suppose we initialize $w_1^{\langle 0 \rangle} = 0.5$, then the Population EM estimate $(\boldsymbol{\theta}^t, w_1^{\langle t \rangle})$ for Model 4 converges to either $(\boldsymbol{\theta}^\star, w_1^\star)$ or $(-\boldsymbol{\theta}^\star, w_2^\star)$ with any initialization $\boldsymbol{\theta}^{\langle 0 \rangle}$ except on the hyperplane $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle = 0$. Furthermore, the convergence speed is linear after some finite number of iterations, i.e., there exists a

finite number T and constant $\rho \in (0,1)$ – depending only on $\boldsymbol{\theta}^\star, w_1^\star$ and initialization $\boldsymbol{\theta}^{\langle 0 \rangle}$– such that

$$\|\boldsymbol{\theta}^{\langle t+1 \rangle} - \kappa\boldsymbol{\theta}^\star\|^2 + |w_1^{\langle t+1 \rangle} - w_1^\star|^2 \leq \rho^{t-T}\left(\|\boldsymbol{\theta}^{\langle T \rangle} - \kappa\boldsymbol{\theta}^\star\|^2 + (w_1^{\langle T \rangle} - w_1^\star)^2\right), \quad \forall t > T,$$

where $\kappa = \text{sgn}(\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle)$.

Theorem 2.4 implies that if we use random initialization for $\boldsymbol{\theta}^{\langle 0 \rangle}$, with probability one, the Population EM estimates converge to the true parameters for Model 4.

The failure of Population EM_3 and success of Population EM_4 can be explained intuitively. Let C_1 and C_2, respectively, denote the true mixture components with parameters $(w_1^\star, \theta^\star)$ and $(w_2^\star, -\theta^\star)$. Due to the symmetry in Population EM_3, we are assured that among the two estimated mixture components, one will have a positive mean, and the other will have a negative mean: call these $\hat{C}_+$ and $\hat{C}_-$, respectively. Assume $\theta^\star > 0$ and $w_1^\star > 0.5$, and consider initializing the Population EM_3 with $\theta^{\langle 0 \rangle} := -\theta^\star$. This initialization incorrectly associates $\hat{C}_-$ with the larger weight $w_1^\star$ instead of the smaller weight $w_2^\star$. This causes, in the E-step of EM, the component $\hat{C}_-$ to become "responsible" for an overly large share of the overall probability mass, and in particular an overly large share of the mass from C_1 (which has a positive mean). Thus, in the M-step of EM, when the mean of the estimated component $\hat{C}_-$ is updated, it is pulled rightward towards $+\infty$. It is possible that this rightward pull would cause the estimated mean of $\hat{C}_-$ to become positive—in which case the roles of $\hat{C}_+$ and $\hat{C}_-$ would switch—but this will not happen as long as $w_1^\star$ is sufficiently bounded away from 1 (but still > 0.5).[1] The result is a bias in the estimation of $\theta^\star$,

[1]When $w_1^\star$ is indeed very close to 1, then almost all of the probability mass of the true distribution comes from C_1, which has positive mean. So, in the M-step discussed above, the rightward pull of the mean of $\hat{C}_-$ may be so strong that the updated mean estimate becomes positive. Since the model enforces that the mean estimates of $\hat{C}_+$ and $\hat{C}_-$ be negations of each other, the roles of $\hat{C}_+$ and $\hat{C}_-$ switch, and now it is $\hat{C}_+$ that becomes associated with the larger mixing weight $w_1^\star$. In this

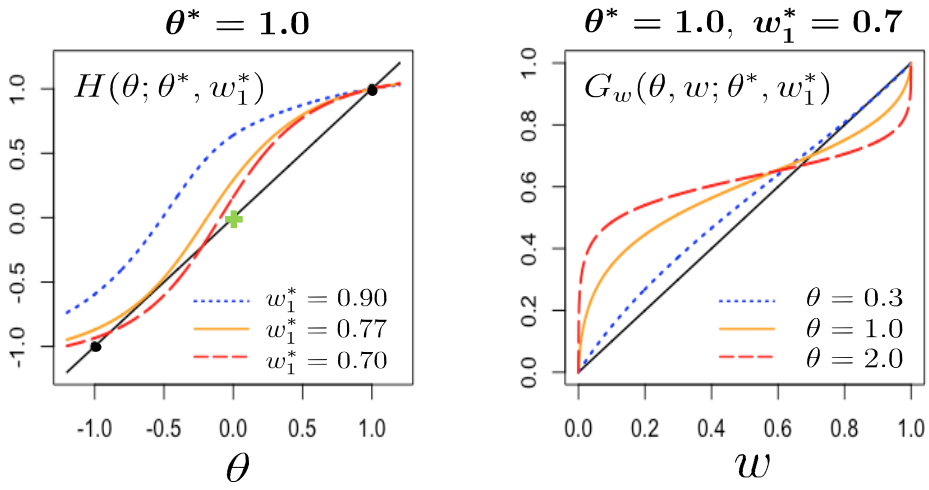

Figure 2.1: Left panel: we show the shape of iterative function $H(\theta; \theta^\star, w_1^\star)$ with $\theta^\star = 1$ and different values of $w_1^\star \in \{0.9, 0.77, 0.7\}$. The green plus + indicates the origin $(0, 0)$ and the black points indicate the correct values $(\theta^\star, \theta^\star)$ and $(-\theta^\star, -\theta^\star)$. We observe that as $w_1^\star$ increases, the number of fixed points goes down from 3 to 2 and finally to 1. Further, when there exists more than one fixed point, there is one stable incorrect fixed point in $(-\theta^\star, 0)$. Right panel: we show the shape of iterative function $G_w(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star)$(defined in Equation (2.25)) with $\theta^\star = 1, w_1^\star = 0.7$ and different values of $\theta \in \{0.3, 1, 2\}$. We observe that as θ increases, G_w becomes from a concave function to a concave-convex function. Further, there are at most three fixed points and there is only one stable fixed point.

thus explaining why the Population EM_3 estimate converges to some $\theta_{\text{wrong}} \in (-\theta^\star, 0)$ when $w_1^\star$ is not too large.

2.3 Main Results for Sample-based EM

Using the results on Population EM presented in the above section, we can now establish consistency of (Sample-based) EM. We focus attention on Model 2 and Model 4, as the same results for Model 1 easily follow as a corollary. The proof of the results are presented in Section 2.5 with more details in Appendix A.2.

---

case, owing to the symmetry assumption, Population EM_3 may be able to successfully converge to $\theta^\star$.

First, we start with Model 2 and state a simple connection between the Population EM and Sample-based EM iterates.

Theorem 2.5. Suppose Population EM and Sample-based EM for Model 2 have the same initial parameters: $\hat{\boldsymbol{\mu}}_1^{\langle 0\rangle} = \boldsymbol{\mu}_1^{\langle 0\rangle}$ and $\hat{\boldsymbol{\mu}}_2^{\langle 0\rangle} = \boldsymbol{\mu}_2^{\langle 0\rangle}$. Then for each iteration $t \geq 0$,

$$\hat{\boldsymbol{\mu}}_1^{\langle t\rangle} \;\to\; \boldsymbol{\mu}_1^{\langle t\rangle} \quad \text{and} \quad \hat{\boldsymbol{\mu}}_2^{\langle t\rangle} \;\to\; \boldsymbol{\mu}_2^{\langle t\rangle} \quad \text{as } n \to \infty\,,$$

where convergence is in probability.

Note that Theorem 2.5 does not necessarily imply that the fixed point of Sample-based EM (when initialized at $(\hat{\boldsymbol{\mu}}_1^{\langle 0\rangle}, \hat{\boldsymbol{\mu}}_2^{\langle 0\rangle}) = (\boldsymbol{\mu}_1^{\langle 0\rangle}, \boldsymbol{\mu}_2^{\langle 0\rangle})$) is the same as that of Population EM. It is conceivable that as $t \to \infty$, the discrepancy between (the iterates of) Sample-based EM and Population EM increases. We show that this is not the case: the fixed points of Sample-based EM indeed converge to the fixed points of Population EM.

Theorem 2.6. Suppose Population EM and Sample-based EM for Model 2 have the same initial parameters: $\hat{\boldsymbol{\mu}}_1^{\langle 0\rangle} = \boldsymbol{\mu}_1^{\langle 0\rangle}$ and $\hat{\boldsymbol{\mu}}_2^{\langle 0\rangle} = \boldsymbol{\mu}_2^{\langle 0\rangle}$. If $\langle \boldsymbol{\mu}_2^{\langle 0\rangle} - \boldsymbol{\mu}_1^{\langle 0\rangle}, \boldsymbol{\theta}^\star \rangle \neq 0$, then

$$\limsup_{t\to\infty} |\hat{\boldsymbol{\mu}}_1^{\langle t\rangle} - \boldsymbol{\mu}_1^{\langle t\rangle}| \;\to\; 0 \quad \text{and} \quad \limsup_{t\to\infty} |\hat{\boldsymbol{\mu}}_2^{\langle t\rangle} - \boldsymbol{\mu}_2^{\langle t\rangle}| \;\to\; 0 \quad \text{as } n \to \infty\,,$$

where convergence is in probability.

Finally, to complete the analysis of EM for the mixtures of two Gaussians, we present the following result that applies to Sample-based EM for Model 4.

Theorem 2.7. Let $(\hat{\boldsymbol{\theta}}^{\langle t\rangle}, \hat{w}_1^{\langle t\rangle})$ be the estimates of Sample-based EM for Model 4. Suppose $\hat{\boldsymbol{\theta}}^{\langle 0\rangle} = \boldsymbol{\theta}^{\langle 0\rangle}, \hat{w}_1^{\langle 0\rangle} = w_1^{\langle 0\rangle} = \frac{1}{2}$ and $\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\theta}^\star \rangle \neq 0$. Then we have

$$\limsup_{t\to\infty} \|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| \;\to\; 0 \quad \text{and} \quad \limsup_{t\to\infty} |\hat{w}_1^{\langle t\rangle} - w_1^{\langle t\rangle}| \;\to\; 0 \quad \text{as } n \to \infty\,,$$

where convergence is in probability.

2.4 Proof for Population EM's results

Due to Lemma 2.3, we assume that $\boldsymbol{\Sigma} = \boldsymbol{I}_d$ for all models for the entire section. Also, due to Lemma 2.1, Population EM_1 is a special case of Population EM_2 with $\boldsymbol{a}^{\langle t \rangle} = \mathbf{0}$ for all $t \geq 0$. Hence, Theorem 2.1 immediately follows Theorem 2.2 with $\boldsymbol{a}^{\langle t \rangle} = \mathbf{0}$. Yet, the analysis for Model 1 in one dimension is a standard analysis for the dynamic system of one update sequence($\{\theta^{\langle t \rangle}\}$ in this case). Further, it plays a key role in the proofs of both Theorem 2.2 and Theorem 2.3 and inspires the proof for Theorem 2.4. Hence, in Subsection 2.4.1, we provide a detailed proof for Model 1 in one dimension followed by the proof for Theorem 2.3.

For Theorem 2.2 and Theorem 2.4, due to the high complexity of the proofs, we present our proofs for the three theorems in four subsections. In Subsection 2.4.2, we prove some properties that shared among Model 2, and 4, and then prove the reduction to the case when $d \leq 2$. Then in Subsection 2.4.3 and Subsection 2.4.4, we prove the theorems respectively for $d \leq 2$.

Notation: We define $\phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}, w)$ be the pdf for the mixture of two Gaussians $w\mathcal{N}(\boldsymbol{\theta}, \boldsymbol{I}) + (1-w)\mathcal{N}(-\boldsymbol{\theta}, \boldsymbol{I})$, i.e., $\phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}, w) = w\phi_d(\boldsymbol{y} - \boldsymbol{\theta}) + (1-w)\phi_d(\boldsymbol{y} + \boldsymbol{\theta})$. Let $\phi^+(y, \theta, w)$ be shorthand for ϕ_d^+ when $d = 1$ and further, $\phi^+(y, \theta)$ be shorthand for $\phi^+(y, \theta, w = 0.5)$. Next, we define the following functions based on the update rules

for our models:

$$G_p(\boldsymbol{a},\boldsymbol{\theta};\boldsymbol{\theta}^\star) \triangleq \int \frac{e^{\langle \boldsymbol{y}-\boldsymbol{a},\boldsymbol{\theta}\rangle}}{e^{\langle \boldsymbol{y}-\boldsymbol{a},\boldsymbol{\theta}\rangle}+e^{-\langle \boldsymbol{y}-\boldsymbol{a},\boldsymbol{\theta}\rangle}}\phi_d^+(\boldsymbol{y},\boldsymbol{\theta}^\star,0.5)\mathrm{d}\boldsymbol{y} \tag{2.22}$$

$$G_\gamma(\boldsymbol{a},\boldsymbol{\theta};\boldsymbol{\theta}^\star) \triangleq \int \frac{e^{\langle \boldsymbol{y}-\boldsymbol{a},\boldsymbol{\theta}\rangle}}{e^{\langle \boldsymbol{y}-\boldsymbol{a},\boldsymbol{\theta}\rangle}+e^{-\langle \boldsymbol{y}-\boldsymbol{a},\boldsymbol{\theta}\rangle}}\boldsymbol{y}\phi_d^+(\boldsymbol{y},\boldsymbol{\theta}^\star,0.5)\mathrm{d}\boldsymbol{y} \tag{2.23}$$

$$H(\boldsymbol{\theta};\boldsymbol{\theta}^\star,w_1^\star) \triangleq \int \frac{w_1^\star e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}-w_2^\star e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}{w_1^\star e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}+w_2^\star e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}\boldsymbol{y}\phi_d^+(\boldsymbol{y},\boldsymbol{\theta}^\star,w_1^\star)\mathrm{d}\boldsymbol{y} \tag{2.24}$$

$$G_w(\boldsymbol{\theta},w_1;\boldsymbol{\theta}^\star,w_1^\star) \triangleq \int \frac{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}+w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}\phi_d^+(\boldsymbol{y},\boldsymbol{\theta}^\star,w_1^\star)\mathrm{d}\boldsymbol{y} \tag{2.25}$$

$$G_\theta(\boldsymbol{\theta},w_1;\boldsymbol{\theta}^\star,w_1^\star) \triangleq \int \frac{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}-w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}+w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}\boldsymbol{y}\phi_d^+(\boldsymbol{y},\boldsymbol{\theta}^\star,w_1^\star)\mathrm{d}\boldsymbol{y} \tag{2.26}$$

where $w_2 = 1 - w_1$ and $w_2^\star = 1 - w_1^\star$. In one dimensional setting, we use lower case $g_p, g_\gamma, g_w, g_\theta$as another representation for $G_p, G_\gamma, G_w, G_\theta$.Specifically, we define

$$g_p(a,\theta;\theta^\star) \triangleq \int \frac{e^{(y-a)\theta}}{e^{(y-a)\theta}+e^{-(y-a)\theta}}\phi^+(y,\theta^\star,0.5)\mathrm{d}y \tag{2.27}$$

$$g_\gamma(a,\theta;\theta^\star) \triangleq \int \frac{e^{(y-a)\theta}}{e^{(y-a)\theta}+e^{-(y-a)\theta}}y\phi^+(y,\theta^\star,0.5)\mathrm{d}y \tag{2.28}$$

$$g_w(\theta,w_1;\theta^\star,w_1^\star) \triangleq \int \frac{w_1 e^{y\theta}}{w_1 e^{y\theta}+w_2 e^{-y\theta}}\phi^+(y,\theta^\star,w_1^\star)\mathrm{d}y \tag{2.29}$$

$$g_\theta(\theta,w_1;\theta^\star,w_1^\star) \triangleq \int \frac{w_1 e^{y\theta}-w_2 e^{-y\theta}}{w_1 e^{y\theta}+w_2 e^{-y\theta}}y\phi^+(y,\theta^\star,w_1^\star)\mathrm{d}y \tag{2.30}$$

Then we define $F : \mathbb{R}^2 \to \mathbb{R}$:

$$\begin{aligned} F(\theta,\theta^\star) &= \int \frac{e^{(y+\theta^\star)\theta}-e^{-(y+\theta^\star)\theta}}{e^{(y+\theta^\star)\theta}+e^{-(y+\theta^\star)\theta}}(y+\theta^\star)\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\mathrm{d}y \\ &= \int (2\mathsf{w}(y+\theta^\star,\theta)-1)(y+\theta^\star)\phi(y)\mathrm{d}y. \end{aligned} \tag{2.31}$$

To understand where this function may appear, note that when $d = 1$,

$$\begin{aligned} g_\gamma(0, \theta, \theta^\star) &= \int \mathsf{w}(y, \theta) y \frac{1}{2} \phi^+(y, \theta^\star) \mathrm{d}y \\ &= \int \mathsf{w}(y, \theta) y \frac{1}{2} \phi(y - \theta^\star) \mathrm{d}y + \int \mathsf{w}(y, \theta) y \frac{1}{2} \phi(y + \theta^\star) \mathrm{d}y \\ &= \int \mathsf{w}(y + \theta^\star, \theta)(y + \theta^\star) \frac{1}{2} \phi(y) \mathrm{d}y + \int \mathsf{w}(y - \theta^\star, \theta)(y - \theta^\star) \frac{1}{2} \phi(y) \mathrm{d}y \\ &= \int \mathsf{w}(y + \theta^\star, \theta)(y + \theta^\star) \frac{1}{2} \phi(y) \mathrm{d}y - \int \mathsf{w}(-y - \theta^\star, \theta)(y + \theta^\star) \frac{1}{2} \phi(y) \mathrm{d}y \\ &= \frac{1}{2} \int \frac{e^{(y+\theta^\star)\theta} - e^{-(y+\theta^\star)\theta}}{e^{(y+\theta^\star)\theta} + e^{-(y+\theta^\star)\theta}} (y + \theta^\star) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \mathrm{d}y = \frac{1}{2} F(\theta, \theta^\star). \end{aligned} \tag{2.32}$$

Similarly, when $d = 1$, we also have

$$H(\theta; \theta^\star, 0.5) \equiv F(\theta, \theta^\star). \tag{2.33}$$

2.4.1 Analysis of Model 1 and Model 3 when $d = 1$

Without loss of generality, we assume that $\theta^\star > 0$ and $\Sigma = 1$. We start with Model 1 providing a detailed proof of the convergence analysis. Then, we analyze Model 3 and prove Theorem 2.3.

2.4.1.1 Model 1

In one dimension, the Population EM iterates of Model 1 follow the following update rule:

$$\theta^{\langle t+1 \rangle} = H(\theta^{\langle t \rangle}; \theta^\star, 0.5) = F(\theta^{\langle t \rangle}, \theta^\star).$$

It is straightforward to show that

- $F(\theta, \theta^\star)$ is a concave function of θ for $\theta \geq 0$ and $F(\theta^\star, \theta^\star) = \theta^\star$.
- $F(\theta, \theta^\star)$ is a convex function of θ for $\theta \leq 0$ and $F(-\theta^\star, \theta^\star) = -\theta^\star$.

- $F(0, \theta^\star) = 0$.

Due to the concavity and convexity of the function, we have

$$\begin{aligned} F(\theta, \theta^\star) - \theta &= \begin{cases} > 0, & \theta \in (-\infty, -\theta^\star) \bigcup (0, \theta^\star) \\ = 0, & \theta = -\theta^\star, 0, \theta^\star \\ < 0, & \theta \in (-\theta^\star, 0) \bigcup (\theta^\star, \infty) \end{cases}, \\ F(\theta, \theta^\star) - \mathrm{sgn}(\theta)\theta^\star &= \begin{cases} < 0, & \theta \in (-\infty, -\theta^\star) \bigcup (0, \theta^\star) \\ > 0, & \theta \in (-\theta^\star, 0) \bigcup (\theta^\star, \infty) \end{cases} \end{aligned} \tag{2.34}$$

Therefore, function F has only three fixed points which are $\{-\theta^\star, 0, \theta^\star\}$. Hence, if we initialize at the fixed points, i.e., $\theta^{\langle 0 \rangle} \in \{-\theta^\star, 0, \theta^\star\}$, then the Population EM iterates will stay at that fixed point, i.e., $\theta^{\langle t \rangle} = \theta^{\langle 0 \rangle}$ for all $t \geq 0$.

Further, from the discussion in Definition 1 and the shape of function $F(\theta, \theta^\star) \equiv H(\theta; \theta^\star, 0.5)$ shown in Equation (2.34), we know $\theta^\star$ and $-\theta^\star$ should be the only two stable fixed points. Hence, when initialization is away from the fixed points, i.e., $\theta^{\langle 0 \rangle} \notin \{-\theta^\star, 0, \theta^\star\}$, we should expect that the iterates $\theta^{\langle t \rangle}$ converges to one of the stable fixed points $\theta^\star$ or $-\theta^\star$. Our next step is to confirm this claim, i.e.,

$$\theta^{\langle t \rangle} \to \mathrm{sgn}(\theta^{\langle 0 \rangle})\theta^\star, \qquad \text{as } t \to \infty. \tag{2.35}$$

Let us show Equation (2.35) when $\theta^{\langle 0 \rangle} > 0$. The case when $\theta^{\langle 0 \rangle} < 0$ follows the same proof due to the fact that $F(\theta, \theta^\star)$ is an odd function of θ. Note that from Equation (2.34), we have

$$(F(\theta, \theta^\star) - \theta)(F(\theta, \theta^\star) + \theta - 2\theta^\star) \ < \ 0, \qquad \forall \theta \in (0, \theta^\star) \bigcup (\theta^\star, \infty)$$

which is equivalent to

$$(F(\theta, \theta^\star) - \theta^\star)^2 \quad < \quad (\theta - \theta^\star)^2, \qquad \forall \theta \in (0, \theta^\star) \bigcup (\theta^\star, \infty). \tag{2.36}$$

Hence, $(\theta^{\langle t \rangle} - \theta^\star)^2$ is a positive decreasing sequence and thus it converges to a non-negative limit. To show Equation (2.35), we just need to confirm that $(\theta^{\langle t \rangle} - \theta^\star)^2$ converges 0 and we prove it by contradiction. Suppose $(\theta^{\langle t \rangle} - \theta^\star)^2$ converges to a non-zero limit, then we know there exists a small enough constant $\epsilon \in (0, \theta^\star/2)$ and large enough iteration $T_0 > 0$ such that

$$\theta^{\langle t \rangle} \in (0, \theta^\star - \epsilon] \bigcup [\theta^\star + \epsilon, \infty), \qquad \forall t \geq T_0.$$

Further, from Equation (2.34), we know that for all $t \geq 0$,

- If $\theta^{\langle t \rangle} \in (0, \theta^\star)$, we have $\theta^{\langle t+1 \rangle} \in (\theta^{\langle t \rangle}, \theta^\star)$
- If $\theta^{\langle t \rangle} \in (\theta^\star, \infty)$, we have $\theta^{\langle t+1 \rangle} \in (\theta^\star, \theta^{\langle t \rangle})$.

Hence, we have

$$\theta^{\langle t+1 \rangle} \in [\min(\theta^\star - 2\epsilon, \theta^{\langle t \rangle}), \theta^\star - \epsilon] \bigcup [\theta^\star + \epsilon, \max(\theta^\star + 2\epsilon, \theta^{\langle t \rangle})], \qquad \forall t \geq T_0.$$

By induction, it is straightforward to show that

$$\theta^{\langle t \rangle} \in [\min(\theta^\star - 2\epsilon, \theta^{\langle T_0 \rangle}), \theta^\star - \epsilon] \bigcup [\theta^\star + \epsilon, \max(\theta^\star + 2\epsilon, \theta^{\langle T_0 \rangle})], \qquad \forall t \geq T_0.$$

Now we have bounded $\theta^{\langle t \rangle}$ in a compact set that excludes 0 and $\theta^\star$. Let us denote this compact set be $\mathcal{C}$. Then note that the function $(F(\theta, \theta^\star) - \theta^\star)^2/(\theta - \theta^\star)^2$ is a well defined and continuous function of θ for θ in the compact set $\mathcal{C}$. Hence, by Equation

(2.36), we know there exists a constant $\rho \in (0, 1)$ such that

$$\frac{(F(\theta, \theta^\star) - \theta^\star)^2}{(\theta - \theta^\star)^2} \leq \rho, \qquad \forall \theta \in \mathcal{C}. \tag{2.37}$$

Hence, we have

$$(\theta^{\langle t+1 \rangle} - \theta^\star)^2 \ < \ \rho(\theta^{\langle t \rangle} - \theta^\star)^2, \qquad \forall t \geq T_0.$$

This implies

$$(\theta^{\langle t \rangle} - \theta^\star)^2 \ < \ \rho^{t-T_0}(\theta^{\langle T_0 \rangle} - \theta^\star)^2, \qquad \forall t \geq T_0,$$

which implies that $(\theta^{\langle t \rangle} - \theta^\star)^2$ converges to 0 and this is a contradiction. Hence, we have shown that $\theta^{\langle t \rangle} \to \theta^\star$ when $\theta^{\langle 0 \rangle} > 0$. The case when $\theta^{\langle 0 \rangle} < 0$ follows the same proof and therefore we have complete the proof for Equation (2.35) and our analysis of for Model 1 under one dimensional setting.

Remark 2.1. From the proof we know that the continuity of function F and its properties in Equation (2.34) are the keys to the proof of convergence to the stable fixed points. We should point out that the contradiction arguments do not guarantee geometric convergence. The geometric convergence can be guaranteed if function F further satisfies that there exists a neighborhood around the stable fixed point such that in the neighborhood, F is differentiable and its derivatives is bounded away from 1. In fact, $F(\theta, \theta^\star)$ for Model 1 has this property and this strategy is also embedded and proved in the analysis for Model 2 and Model 4.

2.4.1.2 Model 3

In one dimension, the Population EM iterates of Model 3 follow the following update rule:

$$\theta^{\langle t+1 \rangle} = H(\theta^{\langle t \rangle}; \theta^\star, w_1^\star),$$

where $w_1^\star > 0.5$. From Definition 1, to prove Theorem 2.3, we need to find a stable fixed points of $H(\theta; \theta^\star, w_1^\star)$ which is not $\theta^\star$ or $-\theta^\star$. Towards this goal, we analyze the shape of $H(\theta; \theta^\star, w_1)$. Note that we have analyzed the shape of H at $w_1 = 0.5$ in Equation (2.34) and conclude that $\pm\theta^\star$ are the only two stable fixed points. Our next step is to compare $H(\theta; \theta^\star, w_1^\star)$ with $F(\theta; \theta^\star) = H(\theta; \theta^\star, 0.5)$ and prove the following lemma.

Lemma 2.2. For all $w_1 \neq 0.5$, we have

$$H(\theta; \theta^\star, w_1) \;>\; H(\theta; \theta^\star, 0.5), \qquad \forall \theta < \theta^\star, \tag{2.38}$$

and for all $w_1 \in [0, 1]$, we have

$$0 \;\leq\; \frac{\partial H(\theta; \theta^\star, w_1)}{\partial \theta} \;\leq\; e^{-\frac{(\theta^\star)^2}{2}} \;<\; 1, \qquad \forall \theta \geq \theta^\star. \tag{2.39}$$

We prove this lemma in Appendix A.1.1.1.

Lemma 2.2 shows that the function curve of $H(\theta; \theta^\star, w_1)$ is strictly above the curve $H(\theta; \theta^\star, 0.5)$ for all $w_1 \neq 0.5$ and $\theta < \theta^\star$, then by Equation (2.34), we have

$$H(\theta; \theta^\star, w_1) - \theta \;>\; H(\theta; \theta^\star, 0.5) - \theta \;\geq\; 0, \qquad \forall w_1 \neq 0.5, \theta \leq -\theta^\star. \tag{2.40}$$

Further, since function $H(\theta; \theta^\star, w_1)$ is continuous with respect to w_1 and from Equation (2.34),

$$H(\theta; \theta^\star, 0.5) < \theta, \qquad \forall \theta \in (-\theta^\star, 0),$$

we know there exists $\delta > 0$ and θ_δ, such that

$$H(\theta_\delta; \theta^\star, w_1) \;<\; \theta_\delta, \quad \forall w_1 \in [0.5, 0.5 + \delta].$$

Hence, with Equation (2.40) and continuity of function $H(\theta;\theta^\star,w_1^\star)-\theta$ with respect to θ, we know when $w_1^\star \in (0.5, 0.5+\delta]$, there exists $\theta_w \in (-\theta^\star, 0)$ (the smallest fixed point) such that

$$H(\theta_w;\theta^\star,w_1^\star) \;=\; \theta_w \quad \text{and} \quad H(\theta;\theta^\star,w_1^\star) \;>\; \theta, \qquad \forall\theta\in(-\infty,\theta_w).$$

Therefore, if we initialize $\theta^{\langle 0\rangle} < -\theta^\star$, then the Population EM iterates $\theta^{\langle t\rangle}$ is an increasing sequence with $\theta^{\langle t\rangle} < \theta_w$ for all $t \geq 0$. Hence, $\theta^{\langle t\rangle}$ should converge to the smallest fixed point of the function H, i.e., $\theta^{\langle t\rangle}$ converges to θ_w. This completes the proof of Theorem 2.3.

In addition, Lemma 2.2 implies the following corollary

Corollary 2.1. For all $w_1^\star \in [0,1]$, $H(\theta;\theta^\star,w_1^\star)$ has only one fixed point (a stable fixed point) in $(0,\infty)$, which is $\theta=\theta^\star$.

2.4.2 Reduction from $d>0$ to the case when $d\leq 2$

In this section, besides the main goal of proving the reduction, we will show some common properties of the Population EM algorithms that are shared among the models. The first property we observe is that the Population EM iterates are rotation invariant for Model 2 and 4.

Lemma 2.3. Let $\boldsymbol{U}\in\mathbb{R}^{d\times d}$ denote a full rank matrix. Let $\boldsymbol{Y}$ be the random variable that follows the true data distribution, i.e., $\boldsymbol{Y}\sim f_2^\star/f_4^\star$ for Model 2/4 respectively. Let $\{(\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{a}^{\langle t\rangle})\}_{t\geq 0}/\{(\boldsymbol{\theta}^{\langle t\rangle},w_1^{\langle t\rangle})\}_{t\geq 0}/\{(\boldsymbol{\theta}^{\langle t\rangle},\sigma^{\langle t\rangle})\}_{t\geq 0}$ be the estimate sequence of Population EM applied on data $\boldsymbol{Y}$ with initialization $(\boldsymbol{\theta}^{\langle 0\rangle},\boldsymbol{a}^{\langle 0\rangle})/(\boldsymbol{\theta}^{\langle 0\rangle},w_1^{\langle 0\rangle})$ for Model 2/4 respectively. Then, we have that $\{(\boldsymbol{U}\boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{U}\boldsymbol{a}^{\langle t\rangle})\}_{t\geq 0}$ or $\{(\boldsymbol{U}\boldsymbol{\theta}^{\langle t\rangle},w_1^{\langle t\rangle})\}_{t\geq 0}$ is the estimate sequence of Population EM applied on data $\boldsymbol{U}\boldsymbol{Y}$ with initialization $(\boldsymbol{U}\boldsymbol{\theta}^{\langle 0\rangle},\boldsymbol{U}\boldsymbol{a}^{\langle 0\rangle})$ or $(\boldsymbol{U}\boldsymbol{\theta}^{\langle 0\rangle},w_1^{\langle 0\rangle})$ for Model 2 or Model 4 respectively.

The proof of Lemma 2.3 is simple by a change of variables in integrations.

With Lemma 2.3, without loss of generality, we assume that $\mathbf{\Sigma} = \boldsymbol{I}$ for all models. Further, Lemma 2.3 implies that at any iteration, we can either directly update the estimates or we first rotate the coordinates, apply one step of Population EM algorithm and then rotate the coordinates back to the original ones. Hence, if our goals are invariant under rotation of the coordinates such as the results stated in Theorem 2.2 and 2.4, we are free to choose any arbitrary coordinate system at any iteration. This flexibility of choosing a desirable coordinate system at each iteration will help us to prove both Theorem 2.2 and 2.4. In fact, using Lemma 2.3, we can prove the following invariance properties of the iterates including the invariance of the sign of $\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^{\star} \rangle$.

Lemma 2.4. For Model 2, we have for all $t \geq 0$,

$$\begin{aligned} \operatorname{sgn}\left(\langle \boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle} \rangle\right) &= \operatorname{sgn}\left(\langle \boldsymbol{a}^{\langle t+1 \rangle}, \boldsymbol{\theta}^{\langle t+1 \rangle} \rangle\right) = \operatorname{sgn}\left(\frac{1}{2} - \mathsf{p}^{\langle t+1 \rangle}\right), \\ \operatorname{sgn}\left(\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^{\star} \rangle\right) &= \operatorname{sgn}\left(\langle \boldsymbol{\theta}^{\langle t+1 \rangle}, \boldsymbol{\theta}^{\star} \rangle\right). \end{aligned}$$

Further, the following holds for any two settings of initializations $(\boldsymbol{a}_{(1)}^{\langle 0 \rangle}, \boldsymbol{\theta}_{(1)}^{\langle 0 \rangle})$ and $(\boldsymbol{a}_{(2)}^{\langle 0 \rangle}, \boldsymbol{\theta}_{(2)}^{\langle 0 \rangle})$.

1. If $\boldsymbol{a}_{(1)}^{\langle 0 \rangle} = -\boldsymbol{a}_{(2)}^{\langle 0 \rangle}$ and $\boldsymbol{\theta}_{(1)}^{\langle 0 \rangle} = \boldsymbol{\theta}_{(2)}^{\langle 0 \rangle}$, then

$$\boldsymbol{a}_{(1)}^{\langle t \rangle} = -\boldsymbol{a}_{(2)}^{\langle t \rangle} \quad \text{and} \quad \boldsymbol{\theta}_{(1)}^{\langle t \rangle} = \boldsymbol{\theta}_{(2)}^{\langle t \rangle}, \qquad \forall t \geq 0.$$

2. If $\boldsymbol{\theta}_{(1)}^{\langle 0 \rangle} = -\boldsymbol{\theta}_{(2)}^{\langle 0 \rangle}$ and $\boldsymbol{a}_{(1)}^{\langle 0 \rangle} = \boldsymbol{a}_{(2)}^{\langle 0 \rangle}$, then

$$\boldsymbol{a}_{(1)}^{\langle t \rangle} = \boldsymbol{a}_{(2)}^{\langle t \rangle} \quad \text{and} \quad \boldsymbol{\theta}_{(1)}^{\langle t \rangle} = -\boldsymbol{\theta}_{(2)}^{\langle t \rangle}, \qquad \forall t \geq 0.$$

Lemma 2.5. For Model 4 with initialization $w_1^{\langle 0 \rangle} = w_2^{\langle 0 \rangle} = 0.5$, we have for all $t \geq 0$,

$$\operatorname{sgn}\left(\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^{\star} \rangle\right) \;=\; \operatorname{sgn}\left(\langle \boldsymbol{\theta}^{\langle t+1 \rangle}, \boldsymbol{\theta}^{\star} \rangle\right) \;=\; \operatorname{sgn}\left(w_1^{\langle t+1 \rangle} - 0.5\right) .$$

Further, for any two settings of initializations $(\boldsymbol{\theta}_{(1)}^{\langle 0 \rangle}, w_{1,(1)}^{\langle 0 \rangle} = 0.5)$ and $(\boldsymbol{\theta}_{(2)}^{\langle 0 \rangle}, w_{1,(2)}^{\langle 0 \rangle} = 0.5)$, if we have $\boldsymbol{\theta}_{(1)}^{\langle 0 \rangle} = -\boldsymbol{\theta}_{(2)}^{\langle 0 \rangle}$, then

$$\boldsymbol{\theta}_{(1)}^{\langle t \rangle} = -\boldsymbol{\theta}_{(2)}^{\langle t \rangle} \quad \text{and} \quad w_{1,(1)}^{\langle t \rangle} + w_{1,(2)}^{\langle t \rangle} = 1 \,, \qquad \forall t \geq 0 \,.$$

We prove above lemmas in Appendix A.1.2.1 and A.1.2.2 respectively.

Lemma 2.4 and Lemma 2.5 implies that it suffices to prove Theorem 2.2 and Theorem 2.4 for the case when $\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^{\star} \rangle \geq 0$ for all $t \geq 0$. Further, we can assume without loss of generality that in Model 2, we have $\langle \boldsymbol{a}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\star} \rangle \geq 0$ and $\mathsf{p}^{\langle t \rangle} \leq 0.5$ for all $t \geq 0$, and in Model 4, we have $w_1^{\langle t \rangle} \geq 0.5$ for all $t \geq 0$. Finally, let us show that the mean estimates of Population EM lie on a subspace with dimension two.

Lemma 2.6. For Model 2, the mean estimates $\boldsymbol{\mu}_1^{\langle t \rangle}$ and $\boldsymbol{\mu}_2^{\langle t \rangle}$ of Population EM lie on the subspace spanned by $\boldsymbol{\theta}^{\star} = \frac{\boldsymbol{\mu}_1^{\star} - \boldsymbol{\mu}_2^{\star}}{2}$ and $\boldsymbol{\theta}^{\langle 0 \rangle} = \frac{\boldsymbol{\mu}_1^{\langle 0 \rangle} - \boldsymbol{\mu}_2^{\langle 0 \rangle}}{2}$. For Model 4, the mean estimates $\boldsymbol{\theta}^{\langle t \rangle}$ of Population EM lie on the subspace spanned by $\boldsymbol{\theta}^{\star}$ and $\boldsymbol{\theta}^{\langle 0 \rangle}$.

Proof. The proof consists of three simple steps. First, rotate the coordinates such that the subspace of the first two coordinates is the same subspace spanned by $\boldsymbol{\theta}^{\star}$ and $\boldsymbol{\theta}^{\langle 0 \rangle}$. Then, using induction, it is straightforward to show that all the mean estimates lie in the subspace of the first two coordinates and thus Lemma 2.3 holds for this new coordinate system. Finally, we can conclude the results of Lemma 2.6 in the original coordinate system by applying Lemma 2.3 and the fact that orthogonality does not change under rotation of the coordinates. □

Due to Lemma 2.6, without loss of generality, we can reduce the d dimensional

problem to an at most two dimensional problem for Model 2 and Model 4.

2.4.3 Proof of Theorem 2.2 when $d \leq 2$

For the two dimensional problem, we can now extend the definition of $\beta^{\langle t \rangle} \in [0, \pi]$ in Equation (2.18) to $\beta^{\langle t \rangle} \in (-\pi, \pi]$. Specifically, for any iteration t, we fix an orthogonal basis $\{\boldsymbol{e}_1^{\langle t \rangle}, \ldots, \boldsymbol{e}_d^{\langle t \rangle}\}$ such that

$$\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{e}_1^{\langle t \rangle} \rangle \;=\; \|\boldsymbol{\theta}^{\langle t \rangle}\| \quad \text{and} \quad \boldsymbol{\theta}^\star \;=\; \langle \boldsymbol{\theta}^\star, \boldsymbol{e}_1^{\langle t \rangle} \rangle \boldsymbol{e}_1^{\langle t \rangle} + \langle \boldsymbol{\theta}^\star, \boldsymbol{e}_2^{\langle t \rangle} \rangle \boldsymbol{e}_2^{\langle t \rangle} \,, \tag{2.41}$$

and define $\beta^{\langle t \rangle} \in (-\pi, \pi]$ be the angle such that

$$\begin{aligned} \cos \beta^{\langle t \rangle} &= \frac{\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^\star \rangle}{\|\boldsymbol{\theta}^{\langle t \rangle}\| \|\boldsymbol{\theta}^\star\|} \,, \\ \sin \beta^{\langle t \rangle} &= \frac{\langle \boldsymbol{\theta}^\star, \boldsymbol{e}_2^{\langle t \rangle} \rangle}{\|\boldsymbol{\theta}^\star\|} \,. \end{aligned}$$

Due to Lemma 2.4, without loss of generality, we assume $|\beta^{\langle t \rangle}| \in [0, \pi/2]$ for all $t \geq 0$ for the rest of this section. Depending on the initialization of $\beta^{\langle 0 \rangle}$, we split the problem into two cases:

(i) $|\beta^{\langle 0 \rangle}| = \pi/2$.

(ii) $|\beta^{\langle 0 \rangle}| \in [0, \pi/2)$.

For case (i), we need to prove the first part of Theorem 2.2, i.e.,

Lemma 2.7. Let $(\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle})_{t \geq 0}$ denote the (re-parameterized) Population EM iterates for Model 2. If $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle = 0$, then

$$(\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle}) \;\to\; (\mathbf{0}, \mathbf{0}) \quad \text{as } t \to \infty \,.$$

We prove this lemma in Appendix A.1.3.5.

Remark 2.2. To show Lemma 2.7, we use a new strategy, part of which is based on prior work in Tseng [2004]. From this approach, one can also show convergence of $(\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})$ for the case $\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\theta}^{\star}\rangle \neq 0$ as well. However, the strategy in Tseng [2004] neither can analyze the convergence speed nor where the estimates converges to directly. Further it is unclear whether it can be generalized to other more complex GMM models.

Now for the rest of the section, we focus on case (ii) when $\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\theta}^{\star}\rangle \neq \mathbf{0}$. From Lemma 2.3, we have the ability to choose arbitrary coordinate system at any iteration as long as our targets are invariant under the rotation of the coordinates. Hence, let us pick one specific sequence of the coordinate systems denoted by $\mathcal{A}$: At iteration t, we rotate the coordinates such that the following holds:

$$\boldsymbol{\theta}^{\langle t\rangle} \;=\; (\|\boldsymbol{\theta}^{\langle t\rangle}\|, 0)^{\top}.$$

To avoid confusion and indicate the coordinate system we are currently using, we use a different notation for $\boldsymbol{\theta}^{\star}, \boldsymbol{\theta}^{\langle t+1\rangle}, \boldsymbol{a}^{\langle t+1\rangle}$ and $\boldsymbol{\gamma}^{\langle t+1\rangle}$. More specifically, when the coordinate system is chosen according to $\boldsymbol{\theta}^{\langle t\rangle}$, we replace $\boldsymbol{\theta}^{\star}$ by $\boldsymbol{\theta}^{\star}_{\langle t\rangle} = (\theta^{\star}_{\langle t\rangle,1}, \theta^{\star}_{\langle t\rangle,2})^{\top}$. For variables at iteration t, we do not change the notation. For variables at $t+1$, we replace $\boldsymbol{\theta}^{\langle t+1\rangle}$ by $\boldsymbol{\theta}^{\langle t+1\rangle}_{\langle t\rangle} = (\theta^{\langle t+1\rangle}_{\langle t\rangle,1}, \theta^{\langle t+1\rangle}_{\langle t\rangle,2})^{\top}$. We define the same notations of $\boldsymbol{a}^{\langle t+1\rangle}_{\langle t\rangle}, a^{\langle t+1\rangle}_{\langle t\rangle,i}$ for $\boldsymbol{a}^{\langle t+1\rangle}$ and $\boldsymbol{\gamma}^{\langle t+1\rangle}_{\langle t\rangle}, \gamma^{\langle t+1\rangle}_{\langle t\rangle,i}$ for $\boldsymbol{\gamma}^{\langle t+1\rangle}$. Finally, we keep the same notation for $\mathsf{p}^{\langle t+1\rangle}$ because $\mathsf{p}^{\langle t+1\rangle}$ is rotation invariant, i.e.,

$$\mathsf{p}^{\langle t+1\rangle} \;=\; G_p(\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^{\star}) \;=\; G_p(\boldsymbol{U}\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{U}\boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{U}\boldsymbol{\theta}^{\star}), \quad \forall \boldsymbol{U}^{\top}\boldsymbol{U} = \boldsymbol{I}.$$

In summary, when the coordinate system is chosen according to $\boldsymbol{\theta}^{\langle t\rangle}$, the update rule

in Equation (2.15) and Equation (2.16) becomes the following:

$$\begin{aligned} \boldsymbol{a}_{\langle t \rangle}^{\langle t+1 \rangle} &= \frac{\boldsymbol{\gamma}_{\langle t \rangle}^{\langle t+1 \rangle}(1 - 2\mathsf{p}^{\langle t+1 \rangle})}{2\mathsf{p}^{\langle t+1 \rangle}(1 - \mathsf{p}^{\langle t+1 \rangle})}, \\ \boldsymbol{\theta}_{\langle t \rangle}^{\langle t+1 \rangle} &= \frac{\boldsymbol{\gamma}_{\langle t \rangle}^{\langle t+1 \rangle}}{2\mathsf{p}^{\langle t+1 \rangle}(1 - \mathsf{p}^{\langle t+1 \rangle})}, \end{aligned} \tag{2.42}$$

where

$$\begin{aligned} \mathsf{p}^{\langle t+1 \rangle} &= \int \mathsf{w}(y_1 - a_1^{\langle t \rangle}, \theta_1^{\langle t \rangle}) \phi^+(y, \theta^\star_{\langle t \rangle, 1}) \\ &= \int \frac{e^{y_1 \|\boldsymbol{\theta}^{\langle t \rangle}\| - a_1^{\langle t \rangle} \|\boldsymbol{\theta}^{\langle t \rangle}\|}}{e^{y_1 \|\boldsymbol{\theta}^{\langle t \rangle}\| - a_1^{\langle t \rangle} \|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{-y_1 \|\boldsymbol{\theta}^{\langle t \rangle}\| + a_1^{\langle t \rangle} \|\boldsymbol{\theta}^{\langle t \rangle}\|}} \frac{1}{2\sqrt{2\pi}} \left(e^{-\frac{(y_1 - \theta^\star_{\langle t \rangle, 1})^2}{2}} + e^{-\frac{(y_1 + \theta^\star_{\langle t \rangle, 1})^2}{2}} \right) \mathrm{d}y_1 \\ &= g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle, 1}). \end{aligned} \tag{2.43}$$

Further, recall that due to Lemma 2.4, we can assume that $\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^\star \rangle > 0$ and $\langle \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{a}^{\langle t \rangle} \rangle \geq 0$ for all $t \geq 0$. Hence, for the rest of the section, we have

$$\theta^\star_{\langle t \rangle, 1} > 0 \quad \text{and} \quad a_1^{\langle t \rangle} \geq 0. \tag{2.44}$$

To prove Theorem 2.2, we show that $\{|\beta^{\langle t \rangle}|\}_{t \geq 0}$ is either a sequence of 0 or a decreasing sequence.

Lemma 2.8. If $\beta^{\langle 0 \rangle} = 0$, then we have $\beta^{\langle t \rangle} = 0$ for all $t \geq 0$. If $\beta^{\langle 0 \rangle} \in (0, \pi/2)$, we have $\pi/2 > \beta^{\langle 0 \rangle} > \beta^{\langle 1 \rangle} > \ldots > \beta^{\langle t \rangle} > \ldots > 0$. If $\beta^{\langle 0 \rangle} \in (-\pi/2, 0)$, we have $-\pi/2 < \beta^{\langle 0 \rangle} < \beta^{\langle 1 \rangle} < \ldots < \beta^{\langle t \rangle} < \ldots < 0$.

Proof. We only prove the claims for $\beta^{\langle 0 \rangle} \in (0, \pi/2)$ because the case when $\beta^{\langle 0 \rangle} \in (-\pi/2, 0)$ follows a similar proof and it is straightforward to check the case when $\beta^{\langle 0 \rangle} = 0$. The strategy of the proof is to use induction to prove that the following three statements hold for $\forall t \geq 0$:

(i) $\beta^{\langle t \rangle} \in (0, \frac{\pi}{2})$.

(ii) $\alpha^{\langle t \rangle} \in (0, \beta^{\langle t \rangle})$.

(iii) $\beta^{\langle t+1 \rangle} = \beta^{\langle t \rangle} - \alpha^{\langle t \rangle} \in (0, \beta^{\langle t \rangle})$.

It is clear that the claim of the lemma holds if (iii) holds for all $t \geq 0$. The inductive argument uses the following chain of arguments for step t:

Claim 1 If (i) holds for t, then (ii) holds for t.

Claim 2 If (i) and (ii) hold for t, then (iii) holds for t.

Claim 3 If (i), (ii), and (iii) hold for t, then (i) holds for $t+1$.

Since (i) holds for $t = 0$ by assumption, it suffices to prove Claims 1–3.

Claim 2 and 3 are trivially true. So we just have to prove Claim 1. Note that the angles $\alpha^{\langle t \rangle}$ and $\beta^{\langle t \rangle}$ for all t are invariant under the rotation of the coordinates. Hence, we apply the sequence of the coordinate systems $\mathcal{A}$ defined previously for the rest of the proof. Then Claim 1 is equivalent to proving that if $\theta^{\star}_{\langle t \rangle, 2} > 0$, then $\alpha^{\langle t \rangle} > 0$ and $\alpha^{\langle t \rangle} < \beta^{\langle t \rangle}$. Therefore, in the rest of the proof, we essentially do these two steps.

1. $\alpha^{\langle t \rangle} > 0$: First note that $\boldsymbol{\theta}^{\langle t+1 \rangle}_{\langle t \rangle}$ and $\boldsymbol{\gamma}^{\langle t+1 \rangle}_{\langle t \rangle}$ are in the same direction. Hence, to

prove that $\alpha^{\langle t \rangle} > 0$ we should show that $\gamma_{\langle t \rangle,2}^{\langle t+1 \rangle} > 0$. We have

$$\begin{aligned}
\gamma_{\langle t \rangle,2}^{\langle t+1 \rangle} &= \int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle}) y_2 \phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}_{\langle t \rangle}^\star) \mathrm{d}\boldsymbol{y} \\
&= \int \mathsf{w}(y_1 - a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|) y_2 \frac{1}{2}(\phi_d(\boldsymbol{y} - \boldsymbol{\theta}_{\langle t \rangle}^\star) + \phi_d(\boldsymbol{y} + \boldsymbol{\theta}_{\langle t \rangle}^\star)) \mathrm{d}\boldsymbol{y} \\
&= \frac{1}{2} \int \mathsf{w}(y_1 - a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|) \phi(y_1 - \theta_{\langle t \rangle,1}^\star) \mathrm{d}y_1 \int y_2 \phi(y_2 - \theta_{\langle t \rangle,2}^\star) \mathrm{d}y_2 \\
&\quad + \frac{1}{2} \int \mathsf{w}(y_1 - a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|) \phi(y_1 + \theta_{\langle t \rangle,1}^\star) \mathrm{d}y_1 \int y_2 \phi(y_2 + \theta_{\langle t \rangle,2}^\star) \mathrm{d}y_2 \\
&= \theta_{\langle t \rangle,2}^\star \int \mathsf{w}(y_1 - a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|) \frac{1}{2}(\phi(y_1 - \theta_{\langle t \rangle,1}^\star) - \phi(y_1 + \theta_{\langle t \rangle,1}^\star)) \mathrm{d}y_1 \\
&= \theta_{\langle t \rangle,2}^\star \int_0^\infty \frac{e^{2y_1\|\boldsymbol{\theta}^{\langle t \rangle}\|} - e^{-2y_1\|\boldsymbol{\theta}^{\langle t \rangle}\|}}{e^{2y_1\|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{2a_1^{\langle t \rangle}\|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{-2y_1\|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{-2a_1^{\langle t \rangle}\|\boldsymbol{\theta}^{\langle t \rangle}\|}} \phi^-(y_1, \theta_{\langle t \rangle,1}^\star) \mathrm{d}y_1 \\
&= \theta_{\langle t \rangle,2}^\star S(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta_{\langle t \rangle,1}^\star), \qquad (2.45)
\end{aligned}$$

where $\phi^-(y, \theta) \triangleq \frac{1}{2}(\phi(y - \theta) - \phi(y + \theta))$ is shorthand for the difference of two Gaussian densities, and $S : \mathbb{R}^3 \to \mathbb{R}$ is defined by

$$\begin{aligned}
&S(a, \theta, \theta^\star) \\
&\quad \triangleq \int_0^\infty \frac{e^{2y\theta} - e^{-2y\theta}}{e^{2y\theta} + e^{-2a\theta} + e^{-2y\theta} + e^{2a\theta}} \cdot \frac{1}{2\sqrt{2\pi}}(e^{-(y-\theta^\star)^2/2} - e^{-(y+\theta^\star)^2/2}) \mathrm{d}y \\
&\quad = \int \mathsf{w}(y - a, \theta) \frac{1}{2}(\phi(y - \theta^\star) - \phi(y + \theta^\star)) \mathrm{d}y \,. \qquad (2.46)
\end{aligned}$$

Hence it is clear that $\theta_{\langle t \rangle,2}^\star > 0$ implies $\alpha^{\langle t \rangle} > 0$ since $S(a, \theta, \theta^\star) > 0$ for all $\theta > 0$ and $\theta^\star > 0$.

2. $\alpha^{\langle t \rangle} < \beta^{\langle t \rangle}$: We just need to show $\alpha^{\langle t \rangle} < \pi/2$ and compare the co-tangent of $\alpha^{\langle t \rangle}$ and $\beta^{\langle t \rangle}$. This means that we have to show $\gamma_{\langle t \rangle,1}^{\langle t+1 \rangle} > 0$ and compare $\frac{\gamma_{\langle t \rangle,1}^{\langle t+1 \rangle}}{\gamma_{\langle t \rangle,2}^{\langle t+1 \rangle}}$ with

$\frac{\theta^\star_{\langle t\rangle,1}}{\theta^\star_{\langle t\rangle,2}}$. We first calculate $\gamma^{\langle t+1\rangle}_{\langle t\rangle,1}$.

$$\begin{aligned}
\gamma^{\langle t+1\rangle}_{\langle t\rangle,1} &= \int \mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}) y_1 \phi^+_d(\boldsymbol{y},\boldsymbol{\theta}^\star_{\langle t\rangle})\mathrm{d}\boldsymbol{y} \\
&= \int \mathsf{w}(y_1-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|) y_1 \phi^+_d(\boldsymbol{y},\boldsymbol{\theta}^\star_{\langle t\rangle})\mathrm{d}\boldsymbol{y} \qquad (2.47)\\
&= \int \mathsf{w}(y_1-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|) y_1 \phi^+(y_1,\theta^\star_{\langle t\rangle,1})\mathrm{d}y_1 \qquad (2.48)\\
&= g_\gamma(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}) \\
&= \int_0^\infty \frac{e^{2y_1\|\boldsymbol{\theta}^{\langle t\rangle}\|}-e^{-2y_1\|\boldsymbol{\theta}^{\langle t\rangle}\|}}{e^{2y_1\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{2a_1^{\langle t\rangle}\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-2y_1\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-2a_1^{\langle t\rangle}\|\boldsymbol{\theta}^{\langle t\rangle}\|}} y_1\phi^+(y_1,\theta^\star_{\langle t\rangle,1})\mathrm{d}y_1, \qquad (2.49)
\end{aligned}$$

where $g_\gamma:\mathbb{R}^3\to\mathbb{R}$ is defined in Equation (2.28). It is clear that $\gamma^{\langle t+1\rangle}_{\langle t\rangle,1}>0$. For comparing the co-tangent of two angles, we need to further simplify $\gamma^{\langle t+1\rangle}_{\langle t\rangle,1}$. We have,

$$\begin{aligned}
\gamma^{\langle t+1\rangle}_{\langle t\rangle,1} &= g_\gamma(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}) \\
&= \int \mathsf{w}(y_1-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|) y_1 \frac{1}{2}(\phi(y_1-\theta^\star_{\langle t\rangle,1})+\phi(y_1+\theta^\star_{\langle t\rangle,1}))\mathrm{d}y_1 \\
&= \theta^\star_{\langle t\rangle,1}\int \mathsf{w}(y_1-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|)\frac{1}{2}(\phi(y_1-\theta^\star_{\langle t\rangle,1})-\phi(y_1+\theta^\star_{\langle t\rangle,1}))\mathrm{d}y_1 \\
&\quad + \int \mathsf{w}(y_1-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|)\frac{1}{2}\left((y_1-\theta^\star_{\langle t\rangle,1})\phi(y_1-\theta^\star_{\langle t\rangle,1})+(y_1+\theta^\star_{\langle t\rangle,1})\phi(y_1+\theta^\star_{\langle t\rangle,1})\right)\mathrm{d}y_1 \\
&= \theta^\star_{\langle t\rangle,1}S(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1})+\int_{-\infty}^{\infty}\mathsf{w}(y_1+\theta^\star_{\langle t\rangle,1}-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|)y_1\frac{1}{2}\phi(y_1)\mathrm{d}y_1 \\
&\quad + \int_{-\infty}^{\infty}\mathsf{w}(y_1-\theta^\star_{\langle t\rangle,1}-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|)y_1\frac{1}{2}\phi(y_1)\mathrm{d}y_1 \\
&= \theta^\star_{\langle t\rangle,1}S(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1})+R(\|\boldsymbol{\theta}^{\langle t\rangle}\|,a_1^{\langle t\rangle}-\theta^\star_{\langle t\rangle,1})+R(\|\boldsymbol{\theta}^{\langle t\rangle}\|,a_1^{\langle t\rangle}+\theta^\star_{\langle t\rangle,1}), \qquad (2.50)
\end{aligned}$$

where $R : \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$\begin{aligned} R(\theta, a) &\triangleq \int_0^{+\infty} \frac{e^{2y\theta} - e^{-2y\theta}}{e^{2y\theta} + e^{2a\theta} + e^{-2y\theta} + e^{-2a\theta}} \frac{1}{2\sqrt{2\pi}} y e^{-y^2/2} \mathrm{d}y \,, \\ &= \int_{-\infty}^{\infty} \mathsf{w}(y - a, \theta) y \frac{1}{2} \phi(y) \mathrm{d}y \,. \end{aligned} \tag{2.51}$$

Employing Equation (2.50) and Equation (2.45) we have

$$\begin{aligned} \cot \alpha^{\langle t \rangle} &= \frac{\theta^{\langle t+1 \rangle}_{\langle t \rangle,1}}{\theta^{\langle t+1 \rangle}_{\langle t \rangle,2}} = \frac{\gamma^{\langle t+1 \rangle}_{\langle t \rangle,1}}{\gamma^{\langle t+1 \rangle}_{\langle t \rangle,2}} \\ &= \frac{\theta^{\star}_{\langle t \rangle,1}}{\theta^{\star}_{\langle t \rangle,2}} + \frac{R(\|\boldsymbol{\theta}^{\langle t \rangle}\|, a_1^{\langle t \rangle} - \theta^{\star}_{\langle t \rangle,1}) + R(\|\boldsymbol{\theta}^{\langle t \rangle}\|, a_1^{\langle t \rangle} + \theta^{\star}_{\langle t \rangle,1})}{\gamma^{\langle t+1 \rangle}_{\langle t \rangle,2}} \\ &> \frac{\theta^{\star}_{\langle t \rangle,1}}{\theta^{\star}_{\langle t \rangle,2}} = \cot \beta^{\langle t \rangle} \,, \end{aligned}$$

where the last inequality is due to the fact that $R(\theta, a) > 0$ for all $\theta > 0$.

□

Lemma 2.8 implies that $\beta^{\langle t \rangle}$ has a limit as iteration t goes to ∞. Our next step is to show that this limit is in fact 0 and prove Equation (2.20), i.e.,

$$\sin(\beta^{\langle t+1 \rangle}) \leq \kappa_\beta^t \cdot \sin(\beta^{\langle 0 \rangle}) \,,$$

where κ_β is some constant in $(0, 1)$ depending only on $\boldsymbol{\theta}^\star$ and initialization $(\boldsymbol{a}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\langle 0 \rangle})$.

Note that Equation (2.20) is invariant under rotation of the coordinates. Hence, we apply the sequence of the coordinate systems $\mathcal{A}$. For notational simplicity we also define $R_s \triangleq R(\|\boldsymbol{\theta}^{\langle t \rangle}\|, a_1^{\langle t \rangle} - \theta^{\star}_{\langle t \rangle,1}) + R(\|\boldsymbol{\theta}^{\langle t \rangle}\|, a_1^{\langle t \rangle} + \theta^{\star}_{\langle t \rangle,1})$ and $S_s \triangleq S(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^{\star}_{\langle t \rangle,1})$.

Since $\boldsymbol{\theta}_{\langle t \rangle}^{\langle t+1 \rangle}$ and $\boldsymbol{\gamma}_{\langle t \rangle}^{\langle t+1 \rangle}$ are in the same direction, we have

$$\begin{aligned}
\cos \beta^{\langle t+1 \rangle} &= \frac{\langle \boldsymbol{\theta}_{\langle t \rangle}^{\langle t+1 \rangle}, \boldsymbol{\theta}_{\langle t \rangle}^{\star} \rangle}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\| \|\boldsymbol{\theta}_{\langle t \rangle}^{\langle t+1 \rangle}\|} = \frac{\langle \boldsymbol{\gamma}_{\langle t \rangle}^{\langle t+1 \rangle}, \boldsymbol{\theta}_{\langle t \rangle}^{\star} \rangle}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\| \|\boldsymbol{\gamma}_{\langle t \rangle}^{\langle t+1 \rangle}\|} \\
&= \frac{\gamma_{\langle t \rangle,1}^{\langle t+1 \rangle} \theta_{\langle t \rangle,1}^{\star} + \gamma_{\langle t \rangle,2}^{\langle t+1 \rangle} \theta_{\langle t \rangle,2}^{\star}}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\| \sqrt{(\gamma_{\langle t \rangle,1}^{\langle t+1 \rangle})^2 + (\gamma_{\langle t \rangle,2}^{\langle t+1 \rangle})^2}} \\
&\overset{\text{(i)}}{=} \frac{(\theta_{\langle t \rangle,1}^{\star})^2 S_s + \theta_{\langle t \rangle,1}^{\star} R_s + (\theta_{\langle t \rangle,2}^{\star})^2 S_s}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\| \sqrt{(\theta_{\langle t \rangle,1}^{\star})^2 S_s^2 + R_s^2 + 2R_s \theta_{\langle t \rangle,1}^{\star} S_s + (\theta_{\langle t \rangle,2}^{\star})^2 S_s^2}} \\
&= \frac{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\|^2 S_s + \theta_{\langle t \rangle,1}^{\star} R_s}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\| \sqrt{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\|^2 S_s^2 + R_s^2 + 2R_s \theta_{\langle t \rangle,1}^{\star} S_s}},
\end{aligned}$$

where Equality (i) is the result of Equation (2.45) and Equation (2.50). Hence it is straightforward to check that

$$\sin \beta^{\langle t+1 \rangle} = \frac{\theta_{\langle t \rangle,2}^{\star} R_s}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\| \sqrt{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\|^2 S_s^2 + R_s^2 + 2R_s \theta_{\langle t \rangle,1}^{\star} S_s}} \leq \frac{\theta_{\langle t \rangle,2}^{\star}}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\|} \frac{R_s}{R_s + \theta_{\langle t \rangle,1}^{\star} S_s}.$$

Note that since $\theta_2^{\langle t \rangle} = 0$, we have

$$\frac{\theta_{\langle t \rangle,2}^{\star}}{\|\boldsymbol{\theta}_{\langle t \rangle}^{\star}\|} = \sin \beta^{\langle t \rangle}.$$

Hence, we have

$$\sin \beta^{\langle t+1 \rangle} \leq \frac{R_s}{R_s + \theta_{\langle t \rangle,1}^{\star} S_s} \sin \beta^{\langle t \rangle}. \tag{2.52}$$

Our goal is to prove that there exists $0 < \kappa_\beta < 1$, such that $R_s/(R_s + \theta_{\langle t \rangle,1}^{\star} S_s) \leq \kappa_\beta$ at every iteration. Toward this goal we will prove that $\theta_{\langle t \rangle,1}^{\star} S_s > 0$. First note that since according to Lemma 2.8 the angle $\beta^{\langle t \rangle}$ is decreasing, $\theta_{\langle t \rangle,1}^{\star}$ is an increasing sequence. Hence, $\theta_{\langle t \rangle,1}^{\star} S_s \geq \theta_{\langle 0 \rangle,1}^{\star} S_s$. Our goal is to show that $S_s > 0$. Note that

$S_s = S(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1})$ is only zero if $\|\boldsymbol{\theta}^{\langle t \rangle}\| = 0$ and can only go to zero if $a_1^{\langle t \rangle} \to \infty$ or $\|\boldsymbol{\theta}^{\langle t \rangle}\| \to \infty$. Hence, if we find a lower bound for $\inf_t \|\boldsymbol{\theta}^{\langle t \rangle}\|$ and prove that we have an upper bound for $\sup_t \|\boldsymbol{a}^{\langle t \rangle}\|$ and $\sup_t \|\boldsymbol{\theta}^{\langle t \rangle}\|$, then we obtain a non-zero lower bound for S_s. The following two lemmas prove our claims:

Lemma 2.9. For any initialization $\boldsymbol{a}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\langle 0 \rangle} \in \mathbb{R}^d$, we have

$$\begin{aligned}
\|\boldsymbol{a}^{\langle t \rangle}\|^2 &\leq \max\left(\|\boldsymbol{a}^{\langle 0 \rangle}\|^2, \frac{2}{\pi} + \frac{\|\boldsymbol{\theta}^\star\|^2}{2}, \frac{16}{9} + \frac{73}{36}\|\boldsymbol{\theta}^\star\|^2\right) \triangleq c_{U,1}^2, \forall t \geq 0, \\
\|\boldsymbol{\theta}^{\langle t \rangle}\|^2 &\leq \max\left(\|\boldsymbol{\theta}^{\langle 0 \rangle}\|^2, \|\boldsymbol{\theta}^\star\|^2 + \frac{1}{4c_{U,2}^2(1-c_{U,2})^2}(1 + \|\boldsymbol{\theta}^\star\|^2)\right) \triangleq c_{U,3}^2, \forall t \geq 0,
\end{aligned}$$

where $c_{U,2} = \frac{1}{4}(1 - \Phi(c_{U,1} + \|\boldsymbol{\theta}^\star\|))$. Hence, $\{\|\boldsymbol{a}^{\langle t \rangle}\|, \|\boldsymbol{\theta}^{\langle t \rangle}\|\}_t$ belong to a compact set.

We prove this lemma in Appendix A.1.3.1, but the fact that the estimates remain bounded should not be surprising for the reader.

Lemma 2.10. Let $\boldsymbol{\theta}^{\langle t \rangle}$ and $\boldsymbol{a}^{\langle t \rangle}$ denote the estimates of Population EM under initialization $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle \neq 0$. There exists a value $c_l > 0$ depending on $\|\boldsymbol{\theta}^\star\|$, $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle$, $\|\boldsymbol{a}^{\langle 0 \rangle}\|$, and $\|\boldsymbol{\theta}^{\langle 0 \rangle}\|$ such that

$$\|\boldsymbol{\theta}^{\langle t \rangle}\| \geq \min(\|\boldsymbol{\theta}^{\langle 0 \rangle}\|, c_l) \triangleq c_{L,1}.$$

We prove this lemma in Appendix A.1.3.2.

Note that according to Lemma 2.9 we know that $\sup_t \|\boldsymbol{a}^{\langle t \rangle}\| \leq c_{U,1}$ and $\sup_t \|\boldsymbol{\theta}^{\langle t \rangle}\| \leq c_{U,3}$. Hence, we define

$$\kappa_\beta = \max_{c_{L,1} \leq \|\boldsymbol{\theta}^{\langle t \rangle}\| \leq c_{U,3}, \|\boldsymbol{a}^{\langle t \rangle}\| \leq c_{U,1}} \frac{R_s}{\theta^\star_{\langle 0 \rangle,1} S(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}) + R_s)} \in (0,1),$$

then Equation (2.52) implies

$$\sin \beta^{\langle t+1 \rangle} \leq \kappa_\beta \sin \beta^{\langle t \rangle}.$$

This proves Equation (2.20) in Theorem 2.2. Note that Equation (2.20) implies that the angle between $\boldsymbol{\theta}^{\langle t \rangle}$ and $\boldsymbol{\theta}^\star$ eventually vanishes. Hence, the convergence behavior of the Population EM estimates $(\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle})$ mainly depends on their behavior under one dimensional setting. Yet, even in one dimensional case, we have two sequences of iterates each of which has the update rule depending on both sequences. Hence, the standard analysis of Model 1 in Section 2.4.1 can not be applied here. As discussed in Section 1.4, it can be hard in general to analyze the convergence behavior for such a coevolving dynamic system. However, in the case of Model 2, the correct fixed points are $(a^{\langle t \rangle}, \theta^{\langle t \rangle}) = (0, \pm\theta^\star)$. Therefore, when Population EM iterates find the global optimum of the maximum likelihood problem, we should expect that $a^{\langle t \rangle}$ converges to 0. Therefore, from Lemma 2.1, Model 1, which has a benign convergence behavior, is not only a special case of Model 2 with $a^{\langle t \rangle} = 0$, it is also the limit to which Model 2 is expected to converge. Hence, our next goal is to prove Equation (2.19), i.e.,

$$\|\boldsymbol{a}^{\langle t+1 \rangle}\|^2 \leq \kappa_a^2 \|\boldsymbol{a}^{\langle t \rangle}\|^2 + \frac{\|\boldsymbol{\theta}^\star\|^2 \sin \beta^{\langle t \rangle}}{4},$$

which implies $a^{\langle t \rangle}$ indeed converges to 0 with Equation (2.20). As before we write $\|\boldsymbol{a}^{\langle t+1 \rangle}\|^2 = (a_{\langle t \rangle,1}^{\langle t+1 \rangle})^2 + (a_{\langle t \rangle,2}^{\langle t+1 \rangle})^2$ and then bound each term separately. According to Equation (2.42), we have

$$a_{\langle t \rangle,1}^{\langle t+1 \rangle} = \frac{g_\gamma(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta_{\langle t \rangle,1}^\star)(1 - 2g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta_{\langle t \rangle,1}^\star))}{g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta_{\langle t \rangle,1}^\star)(1 - g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta_{\langle t \rangle,1}^\star))} \overset{\text{(i)}}{\leq} \kappa_a a_1^{\langle t \rangle} \leq \kappa_a \|\boldsymbol{a}^{\langle t \rangle}\| \quad (2.53)$$

where Inequality (i) is due to the following lemma:

Lemma 2.11. For any $x_{\theta^\star} \geq 0$, there exists a constant $\kappa_a \in (0,1)$ only depending on $x_{\theta^\star}$ and continuous for $x_{\theta^\star} > 0$ such that

$$\frac{g_\gamma(a,\theta,x_{\theta^\star})\left(1-2g_p(a,\theta,x_{\theta^\star})\right)}{2g_p(a,\theta,x_{\theta^\star})\left(1-g_p(a,\theta,x_{\theta^\star})\right)} \leq \kappa_a a, \quad \forall a \geq 0, \theta > 0. \tag{2.54}$$

We prove this lemma in Appendix A.1.3.3.

Our next step is to establish the convergence of $a_{\langle t\rangle,2}^{\langle t+1\rangle}$. From Equation (2.42) and Equation (2.45) we have:

$$\begin{aligned} a_{\langle t\rangle,2}^{\langle t+1\rangle} &= \frac{\gamma_{\langle t\rangle,2}^{\langle t+1\rangle}(1-2g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star))}{2g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star)(1-g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star))} \\ &= \theta_{\langle t\rangle,2}^\star \frac{S(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star)(1-2g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star))}{2g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star)(1-g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star))}. \end{aligned}$$

And according to Equation (2.43), we have

$$\begin{aligned} g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star) &= \int \mathsf{w}(y-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|)\phi^+(y,\theta_{\langle t\rangle,1}^\star)\mathrm{d}y \\ &= \int_{y=0}^{\infty}(\mathsf{w}(y-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|)+\mathsf{w}(-y-a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|))\phi^+(y,\theta_{\langle t\rangle,1}^\star)\mathrm{d}y \\ &= \int_{y=0}^{\infty}\frac{e^{2y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+2e^{-2a_1^{\langle t\rangle}\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-2y\|\boldsymbol{\theta}^{\langle t\rangle}\|}}{e^{2y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{2a_1^{\langle t\rangle}\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-2y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-2a_1^{\langle t\rangle}\|\boldsymbol{\theta}^{\langle t\rangle}\|}}\phi^+(y,\theta_{\langle t\rangle,1}^\star)\mathrm{d}y \\ &> S(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta_{\langle t\rangle,1}^\star) \geq 0. \end{aligned} \tag{2.55}$$

Hence, we have

$$a_{\langle t\rangle,2}^{\langle t+1\rangle} \leq \frac{\theta_{\langle t\rangle,2}^\star}{2} = \frac{\|\theta^\star\|\sin\beta^{\langle t\rangle}}{2}. \tag{2.56}$$

Combining Equation (2.53) and Equation (2.56) establishes Equation (2.19) in our Theorem 2.2.

From Equation (2.20) and Equation (2.19), we can confirm that Model 1 is indeed the limit where Model 2 converges. Note that Equation (2.36) proves the convergence of $\theta^{\langle t \rangle}$ for Model 1. Hence, our final goal is to prove Equation (2.21), i.e.,

$$\left\| \boldsymbol{\theta}^{\langle t+1 \rangle} - \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\star} \rangle) \boldsymbol{\theta}^{\star} \right\|^2 \;\leq\; \kappa_{\theta}^2 \cdot \left\| \boldsymbol{\theta}^{\langle t \rangle} - \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\star} \rangle) \boldsymbol{\theta}^{\star} \right\|^2 + c_{\theta} \cdot \|\boldsymbol{a}^{\langle t \rangle}\| \quad \forall t > T_0 \,,$$

which can be considered as a perturbation and stronger version of Equation (2.36).

It is straightforward to use Equation (2.19) and Equation (2.20) to show that for every $\delta_a > 0$, there exists a value of T_{δ_a} such that for every $t > T_{\delta_a}$, $\|\boldsymbol{a}^{\langle t \rangle}\| \leq \delta_a$. For the moment suppose that the following claim is true: there exists $\delta_a > 0, \kappa_{\theta}, c_{\theta}$ only depending on $\boldsymbol{\theta}^{\star}$ and the initialization $(\boldsymbol{a}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\langle 0 \rangle})$, such that if $\|\boldsymbol{a}^{\langle t \rangle}\| \leq \delta_a$ for some t, then the next iteration $\boldsymbol{\theta}^{\langle t+1 \rangle}$ satisfies the following equation:

$$\|\boldsymbol{\theta}^{\langle t+1 \rangle} - \boldsymbol{\theta}^{\star}\|^2 \;\;\leq\;\; \kappa_{\theta}^2 \|\boldsymbol{\theta}^{\langle t \rangle} - \boldsymbol{\theta}^{\star}\|^2 + c_{\theta} \|\boldsymbol{a}^{\langle t \rangle}\|.$$

If we combine this claim with Equation (2.19) and Equation (2.20), we obtain Equation (2.21). Hence, the problem reduces to proving the above claim.

Note that in Lemma 2.9, Lemma 2.10 and Lemma 2.8, we have

$$\|\boldsymbol{a}^{\langle t \rangle}\| \in [0, c_{U,1}], \|\boldsymbol{\theta}^{\langle t \rangle}\| \in [c_{L,1}, c_{U,3}], \theta^{\star}_{\langle t \rangle, 1} \in [\theta^{\star}_{\langle 0 \rangle, 1}, \|\boldsymbol{\theta}^{\star}\|], \forall t \geq 0.$$

Therefore, it is again straightforward to see that the following lemma implies our claim:

Lemma 2.12. For any $\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle}, \boldsymbol{\theta}^{\star} \in \mathbb{R}^2$, if $\|\boldsymbol{a}^{\langle t \rangle}\| \in [0, U_a], \|\boldsymbol{\theta}^{\langle t \rangle}\| \in [L_{\theta}, U_{\theta}], \frac{\langle \boldsymbol{\theta}^{\star}, \boldsymbol{\theta}^{\langle t \rangle} \rangle}{\|\boldsymbol{\theta}^{\langle t \rangle}\|} \in [L_{\theta^{\star}}, \|\boldsymbol{\theta}^{\star}\|], \forall t \geq 0$, where $L_{\theta} > 0, L_{\theta^{\star}} > 0$ then there exists $\delta_a \in (0, \min\{L_{\theta^{\star}}, 1\}]; \kappa_{\theta} \in (0, 1); c_{\theta} > 0$ such that $\forall \|\boldsymbol{a}^{\langle t \rangle}\| \in [0, \delta_a], \|\boldsymbol{\theta}^{\langle t \rangle}\| \in [L_{\theta}, U_{\theta}], \frac{\langle \boldsymbol{\theta}^{\star}, \boldsymbol{\theta}^{\langle t \rangle} \rangle}{\|\boldsymbol{\theta}^{\langle t \rangle}\|} \in [L_{\theta^{\star}}, \|\boldsymbol{\theta}^{\star}\|]$, the

next iteration $\boldsymbol{\theta}^{\langle t+1\rangle}$ satisfying

$$\|\boldsymbol{\theta}^{\langle t+1\rangle} - \boldsymbol{\theta}^{\star}\|^2 \;\; \leq \;\; \kappa_{\theta}^2 \|\boldsymbol{\theta}^{\langle t\rangle} - \boldsymbol{\theta}^{\star}\|^2 + c_{\theta}\|\boldsymbol{a}^{\langle t\rangle}\|,$$

where δ_a, κ_θ and c_θ are functions of only $U_a, L_\theta, U_\theta, L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|$.

Proof. To prove this lemma, note that its statement is rotation invariant, therefore we apply the sequence of coordinate systems $\mathcal{A}$. Our strategy of proving this lemma is to prove the following two claims :

1. There exists $\kappa_s \in (0,1)$ such that $|\theta^{\langle t+1\rangle}_{\langle t\rangle,2} - \theta^{\star}_{\langle t\rangle,2}| \leq \kappa_s |\theta^{\star}_{\langle t\rangle,2}|$.

2. There exists $\kappa'_\theta \in (0,1)$ and $\delta_a > 0$ such that if $\|\boldsymbol{a}^{\langle t\rangle}\| \leq \delta_a$, then

$$|\theta^{\langle t+1\rangle}_{\langle t\rangle,1} - \theta^{\star}_{\langle t\rangle,1}| \leq \kappa'_\theta \left| \|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta^{\star}_{\langle t\rangle,1} \right| + (16\|\boldsymbol{\theta}^\star\| + 6)\|\boldsymbol{a}^{\langle t\rangle}\|.$$

We will then combine the above two claims to obtain Lemma 2.12.

1. Proof of $|\theta^{\langle t+1\rangle}_{\langle t\rangle,2} - \theta^{\star}_{\langle t\rangle,2}| \leq \kappa_s |\theta^{\star}_{\langle t\rangle,2}|$:

   To prove our first claim, first note that according to Equation (2.42) and Equation (2.45) we have

$$\begin{aligned}
\theta^{\langle t+1\rangle}_{\langle t\rangle,2} &= \frac{\gamma^{\langle t+1\rangle}_{\langle t\rangle,2}}{2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1}))} \\
&= \theta^{\star}_{\langle t\rangle,2} \frac{S(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})}{2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1}))}
\end{aligned}$$

Hence

$$\begin{aligned}|\theta_{\langle t\rangle,2}^{\langle t+1\rangle} - \theta_{\langle t\rangle,2}^{\star}| &= |\theta_{\langle t\rangle,2}^{\star}| \left(1 - \frac{S(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})}{2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}))}\right) \\ &\leq |\theta_{\langle t\rangle,2}^{\star}|(1 - 2S(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}))\end{aligned}$$

By definition of function S in Equation (2.46), it is straightforward to conclude that $S(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})$ is only zero if $\|\boldsymbol{\theta}^{\langle t\rangle}\| = 0$ and can only go to zero if $a_1^{\langle t\rangle} \to \infty$ or $\|\boldsymbol{\theta}^{\langle t\rangle}\| \to \infty$. Therefore, combined with the continuity of S, we conclude that

$$\kappa_s \triangleq \sup_{a_1^{\langle t\rangle}\in[0,U_a], \|\boldsymbol{\theta}^{\langle t\rangle}\|\in[L_\theta,U_\theta], \theta_{\langle t\rangle,1}^{\star}\in[L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]} 1 - 2S(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) < 1, \quad (2.57)$$

where κ_s only depends on $U_a, L_\theta, U_\theta, L_{\theta^\star}$ and $\|\boldsymbol{\theta}^\star\|$. Hence,

$$|\theta_{\langle t\rangle,2}^{\langle t+1\rangle} - \theta_{\langle t\rangle,2}^{\star}| \leq \kappa_s |\theta_{\langle t\rangle,2}^{\star}|. \quad (2.58)$$

2. Proof of $|\theta_{\langle t\rangle,1}^{\langle t+1\rangle} - \theta_{\langle t\rangle,1}^{\star}| \leq \kappa_\theta' |\|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta_{\langle t\rangle,1}^{\star}| + (16\|\theta\| + 6)\|\boldsymbol{a}^{\langle t\rangle}\|$:

Note that

$$\begin{aligned}
|\theta_{\langle t\rangle,1}^{\langle t+1\rangle} - \theta_{\langle t\rangle,1}^{\star}| &= \left| \frac{g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})}{2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}))} - \theta_{\langle t\rangle,1}^{\star} \right| \\
&= \left| \frac{2g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - \theta_{\langle t\rangle,1}^{\star} + \theta_{\langle t\rangle,1}^{\star}\left(2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - 1\right)^2}{4g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}))} \right| \\
&\leq \left| \frac{2g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - \theta_{\langle t\rangle,1}^{\star}}{4g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}))} \right| \\
&\quad + \left| \frac{\theta_{\langle t\rangle,1}^{\star}\left(2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - 1\right)^2}{4g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}))} \right|
\end{aligned} \tag{2.59}$$

Furthermore in Equation (A.33) in the proof of Lemma 2.11, we proved that

$$g_\gamma(a, \theta, x_{\theta^\star}) = a \cdot g_p(a, \theta, x_{\theta^\star}) - \frac{1}{2}a + \frac{1}{4}(F(\theta, a + x_{\theta^\star}) + F(\theta, x_{\theta^\star} - a)).$$

Hence, we have

$$\begin{aligned}
&|2g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - \theta_{\langle t\rangle,1}^{\star}| \\
= &\left| a_1^{\langle t\rangle}(2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - 1) + \frac{1}{2}(F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star} - a_1^{\langle t\rangle}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})) \right. \\
&\left. + \frac{1}{2}(F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star} + a_1^{\langle t\rangle}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})) + (F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - \theta_{\langle t\rangle,1}^{\star}) \right|.
\end{aligned} \tag{2.60}$$

Combining Equation (2.59) and Equation (2.60), we conclude that in order to obtain an upper bound for $|\theta_{\langle t\rangle,1}^{\langle t+1\rangle} - \theta_{\langle t\rangle,1}^{\star}|$ we have to find the following bounds:

(a) Obtain an upper bound for $|2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star}) - 1|$.

(b) Obtain an upper bound for $|F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, x_{\theta^\star}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta_{\langle t\rangle,1}^{\star})|$ for all $\theta_{\langle t\rangle,1}^{\star} \in$

$[L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$ and $|x_{\theta^\star} - \theta^\star_{\langle t\rangle,1}| \leq L_{\theta^\star}$.

(c) Obtain an upper bound for $|F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) - \theta^\star_{\langle t\rangle,1}|$ for all $\theta^\star_{\langle t\rangle,1} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$ and $\|\boldsymbol{\theta}^{\langle t\rangle}\| \in [L_\theta, U_\theta]$

(d) Obtain a lower bound for $4g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))$.

We summarize our strategy for bounding each of these terms below:

(a) Upper bound for $|2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) - 1|$: It is straightforward to confirm

$$2g_p(0, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) = \frac{1}{2}.$$

Hence, we have to calculate $|2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) - 2g_p(0, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})|$. According to mean value theorem

$$\begin{aligned}
&\left|g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) - g_p(0, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})\right| \\
&\qquad = \left|\frac{\partial g_p(a, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})}{\partial a}\Big|_{a=\xi}\right| (a_1^{\langle t\rangle}), \qquad (2.61)
\end{aligned}$$

where $\xi \in [0, a_1^{\langle t\rangle}]$. Therefore we only need to bound $\left|\frac{\partial g_p(a, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})}{\partial a}\right|$. Note that

$$\begin{aligned}
&\left|\frac{\partial g_p(a, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))}{\partial a}\Big|_{a=0}\right| \\
&= \int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\| - a\|\boldsymbol{\theta}^{\langle t\rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\| + a\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2} \phi^+(y, \theta^\star_{\langle t\rangle,1})\mathrm{d}y\Big|_{a=0} \\
&= \int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2} \phi^+(y, \theta^\star_{\langle t\rangle,1})\mathrm{d}y \\
&= \int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2} \phi(y - \theta^\star_{\langle t\rangle,1})\mathrm{d}y. \qquad (2.62)
\end{aligned}$$

Next we show that $\left|\frac{\partial g_p(a,\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}))}{\partial a}\big|_{a=0}\right|$ is a decreasing function of $\theta^\star_{\langle t\rangle,1}$ and hence can be upper bounded by $\left|\frac{\partial g_p(a,\|\boldsymbol{\theta}^{\langle t\rangle}\|,0))}{\partial a}\big|_{a=0}\right|$:

$$\begin{aligned}
&\frac{\partial \int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2}\phi(y-\theta^\star_{\langle t\rangle,1})\mathrm{d}y}{\partial \theta^\star_{\langle t\rangle,1}} \\
= &\int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2}(y-\theta^\star_{\langle t\rangle,1})\phi(y-\theta^\star_{\langle t\rangle,1})\mathrm{d}y \\
= &-\int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2}d\phi(y-\theta^\star_{\langle t\rangle,1}) \\
= &-\int \frac{4\|\boldsymbol{\theta}^{\langle t\rangle}\|^2(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|}-e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})^3}\phi(y-\theta^\star_{\langle t\rangle,1})\mathrm{d}y \leq 0
\end{aligned}$$

Hence,

$$\begin{aligned}
\left|\frac{\partial g_p(a,\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}))}{\partial a}\big|_{a=0}\right| &\leq \int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\|}+e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2}\phi(y)\mathrm{d}y \\
&\leq \int_0^\infty 4\|\boldsymbol{\theta}^{\langle t\rangle}\|e^{-2y\|\boldsymbol{\theta}^{\langle t\rangle}\|}\phi(y)\mathrm{d}y \leq \sqrt{\frac{2}{\pi}}.
\end{aligned} \tag{2.63}$$

Our next step is to show that there exists $\delta_1 > 0$ is a function of only $L_\theta, U_\theta, L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|$ such that

$$\sup_{a_1^{\langle t\rangle}\in[0,\delta_1],\|\boldsymbol{\theta}^{\langle t\rangle}\|\in[L_\theta,U_\theta],\theta^\star_{\langle t\rangle,1}\in[L_{\theta^\star},\|\boldsymbol{\theta}^\star\|]}\left|\frac{\partial g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1})}{\partial a_1^{\langle t\rangle}}\right| \leq 1. \tag{2.64}$$

This is a simple proof by contradiction. Since we have already done similar arguments in the proof of Lemma A.3, for the sake of brevity we skip this argument. By combining Equation (2.61) and Equation (2.64) we conclude

for all $a_1^{\langle t \rangle} \in [0, \delta_1], \|\boldsymbol{\theta}^{\langle t \rangle}\| \in [L_\theta, U_\theta], \theta^\star_{\langle t \rangle,1} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$,

$$|1 - 2g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1})| \leq 2a_1^{\langle t \rangle}. \tag{2.65}$$

(b) Upper bound for $|F(\|\boldsymbol{\theta}^{\langle t \rangle}\|, x_{\theta^\star}) - F(\|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1})|$: Again by employing the mean value theorem, we conclude that we have to bound $\frac{\partial F(\|\boldsymbol{\theta}^{\langle t \rangle}\|, x_{\theta^\star})}{\partial x_{\theta^\star}}$ in a neighborhood of $x_{\theta^\star} = \theta^\star_{\langle t \rangle,1}$ for all $\|\boldsymbol{\theta}^{\langle t \rangle}\| \in [L_\theta, U_\theta], \theta^\star_{\langle t \rangle,1} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$. Note that, $\forall x_{\theta^\star} \geq 0$

$$\begin{aligned}
\left|\frac{\partial F(\|\boldsymbol{\theta}^{\langle t \rangle}\|, x_{\theta^\star})}{\partial x_{\theta^\star}}\right| &= \left|\int (2\mathsf{w}(y, \|\boldsymbol{\theta}^{\langle t \rangle}\|) - 1) y (y - x_{\theta^\star}) \phi(y - x_{\theta^\star}) \mathrm{d}y\right| \\
&= \left|\int (2\mathsf{w}(y, \|\boldsymbol{\theta}^{\langle t \rangle}\|) - 1) \{(y - x_{\theta^\star})^2 + x_{\theta^\star}(y - x_{\theta^\star})\} \phi(y - x_{\theta^\star}) \mathrm{d}y\right| \\
&= \left| x_{\theta^\star} \int \frac{e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} - e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}}{e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}} (y - x_{\theta^\star}) \phi(y - x_{\theta^\star}) \mathrm{d}y \right. \\
&\quad \left. + \int \frac{e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} - e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}}{e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}} (y - x_{\theta^\star})^2 \phi(y - x_{\theta^\star}) \mathrm{d}y \right| \\
&\overset{\text{(i)}}{<} \left| x_{\theta^\star} \int \frac{4\|\boldsymbol{\theta}^{\langle t \rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|})^2} \phi(y - x_{\theta^\star}) \mathrm{d}y \right| + 1 \\
&\overset{\text{(ii)}}{=} 2x_{\theta^\star} \left| \frac{\partial g_p(a, \|\boldsymbol{\theta}^{\langle t \rangle}\|, x_{\theta^\star})}{\partial a}\Big|_{a=0} \right| + 1,
\end{aligned}$$

where to obtain Inequality (i) we used integration by parts and also the fact that $\frac{e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} - e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}}{e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}} < 1$. To see why (ii) holds, one may check Equation (2.62). By employing Equation (2.63), we then conclude that

$$\left|\frac{\partial F(\|\boldsymbol{\theta}^{\langle t \rangle}\|, x_{\theta^\star})}{\partial x_{\theta^\star}}\right| < x_{\theta^\star} \frac{4}{\sqrt{2\pi}} + 1 \leq 4\|\boldsymbol{\theta}^\star\| + 1, \forall x_{\theta^\star} \in [0, 2\|\boldsymbol{\theta}^\star\|].$$

Therefore, using mean value theorem, we have $\forall |x_{\theta^\star} - \theta^\star_{\langle t \rangle,1}| \leq L_{\theta^\star}, \theta^\star_{\langle t \rangle,1} \in$

$[L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$,

$$|F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, x_{\theta^\star}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})| \leq (4\|\boldsymbol{\theta}^\star\| + 1)|x_{\theta^\star} - \theta^\star_{\langle t\rangle,1}|.$$

Hence, we have $\forall a_1^{\langle t\rangle} \in [0, L_{\theta^\star}], \theta^\star_{\langle t\rangle,1} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$,

$$\begin{aligned}&\left|\frac{1}{2}(F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1} - a_1^{\langle t\rangle}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))\right.\\ &\left.+\frac{1}{2}(F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1} + a_1^{\langle t\rangle}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))\right| \;\leq\; (4\|\boldsymbol{\theta}^\star\| + 1)a_1^{\langle t\rangle}.\end{aligned} \tag{2.66}$$

(c) Upper bound for $|F(\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) - \theta^\star_{\langle t\rangle,1}|$:

Because the proof of this part has many algebraic steps we postpone it to Appendix A.1.3.4.

Lemma 2.13. Given $\theta \in [L_\theta, U_\theta], x_{\theta^\star} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$ where $0 < L_\theta \leq L_{\theta^\star} \leq \|\boldsymbol{\theta}^\star\| \leq U_\theta < \infty$, there exists $\kappa''_\theta \in (0, 1)$ is a function of only $L_\theta, U_\theta, L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|$ such that

$$|F(\theta, x_{\theta^\star}) - x_{\theta^\star}| \leq \kappa''_\theta|\theta - x_{\theta^\star}|, \forall\theta \in [L_\theta, U_\theta], x_{\theta^\star} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|].$$

(d) Lower bound for $4g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))$: Note that

$$\frac{1}{4g_p(0, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})(1 - g_p(0, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))} - 1 = 0.$$

Let $\epsilon_p = \min\{\frac{1-\kappa''_\theta}{2\kappa''_\theta}, 1\}$ (This choice will become clear later in the proof). Using contradiction arguments similar to the ones employed in the proof of Lemma A.3, it is straight forward to see that there exists $\delta_2 > 0$ only

depending on $L_\theta, U_\theta, L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|$ such that

$$\sup_{\substack{a_1^{\langle t\rangle}\in[0,\delta_2],\|\boldsymbol{\theta}^{\langle t\rangle}\|\in[L_\theta,U_\theta],\\ \theta^\star_{\langle t\rangle,1}\in[L_{\theta^\star},\|\boldsymbol{\theta}^\star\|]}} \frac{1}{4g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1})(1-g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}))} - 1 \leq \epsilon_p. \tag{2.67}$$

Now combining Equation (2.59), Equation (2.60), Equation (2.65), Equation (2.66), Equation (2.67) and Lemma 2.13 we conclude that for all $\|\boldsymbol{\theta}^{\langle t\rangle}\| \in [L_\theta, U_\theta]$, $a_1^{\langle t\rangle} \in [0, \min\{\delta_1, L_{\theta^\star}, 1\}]$ and $\theta^\star_{\langle t\rangle,1} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$,

$$\begin{aligned}
& |2g_\gamma(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}) - \theta^\star_{\langle t\rangle,1}| \\
\leq\ & |a_1^{\langle t\rangle}(2g_p(a_1^{\langle t\rangle},\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1})-1)| + |\frac{1}{2}(F(\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}-a_1^{\langle t\rangle}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}))| \\
& + |\frac{1}{2}(F(\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}+a_1^{\langle t\rangle}) - F(\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}))| + |F(\|\boldsymbol{\theta}^{\langle t\rangle}\|,\theta^\star_{\langle t\rangle,1}) - \theta^\star_{\langle t\rangle,1}| \\
\leq\ & 2(a_1^{\langle t\rangle})^2 + (4\|\boldsymbol{\theta}^\star\|+1)a_1^{\langle t\rangle} + \kappa''_\theta|\|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta^\star_{\langle t\rangle,1}| \\
\leq\ & (4\|\boldsymbol{\theta}^\star\|+3)a_1^{\langle t\rangle} + \kappa''_\theta|\|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta^\star_{\langle t\rangle,1}|.
\end{aligned}$$

Hence together with Equation (2.65) again and Equation (2.67) in Equation

(2.59), we have $\forall a_1^{\langle t \rangle} \in [0, \min\{\delta_1, L_{\theta^\star}, 1, \delta_2\}], \|\boldsymbol{\theta}^{\langle t \rangle}\| \in [L_\theta, U_\theta], \theta^\star_{\langle t \rangle,1} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$

$$
\begin{aligned}
|\theta^{\langle t+1 \rangle}_{\langle t \rangle,1} - \theta^\star_{\langle t \rangle,1}| &\leq \left| \frac{2g_\gamma(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}) - \theta^\star_{\langle t \rangle,1}}{4g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1})(1 - g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}))} \right| \\
&\quad + \left| \frac{\theta^\star_{\langle t \rangle,1}(2g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}) - 1)^2}{4g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1})(1 - g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}))} \right| \\
&\leq (1+\epsilon_p) \left(\left| 2g_\gamma(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}) - \theta^\star_{\langle t \rangle,1} \right| + \theta^\star_{\langle t \rangle,1} \left(2g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}) - 1 \right)^2 \right) \\
&\leq (1+\epsilon_p)((4\|\boldsymbol{\theta}^\star\| + 3)a_1^{\langle t \rangle} + \kappa''_\theta | \|\boldsymbol{\theta}^{\langle t \rangle}\| - \theta^\star_{\langle t \rangle,1}| + 4\|\boldsymbol{\theta}^\star\| a_1^{\langle t \rangle}) \\
&\leq 2(8\|\boldsymbol{\theta}^\star\| + 3)a_1^{\langle t \rangle} + \left(\frac{1-\kappa''_\theta}{2\kappa''_\theta} + 1 \right) \kappa''_\theta | \|\boldsymbol{\theta}^{\langle t \rangle}\| - \theta^\star_{\langle t \rangle,1}| \\
&\leq (16\|\boldsymbol{\theta}^\star\| + 6)\|\boldsymbol{a}^{\langle t \rangle}\| + \frac{1+\kappa''_\theta}{2} | \|\boldsymbol{\theta}^{\langle t \rangle}\| - \theta^\star_{\langle t \rangle,1}|.
\end{aligned}
$$

In summary, if we set $\delta_a \triangleq \min\{\delta_1, L_{\theta^\star}, 1, \delta_2\} > 0$ and $\kappa'_\theta \triangleq \frac{1+\kappa''_\theta}{2} < 1$, we have $\forall a_1^{\langle t \rangle} \leq \|\boldsymbol{a}^{\langle t \rangle}\| \in [0, \delta_a], \|\boldsymbol{\theta}^{\langle t \rangle}\| \in [L_\theta, U_\theta], \theta^\star_{\langle t \rangle,1} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$,

$$
|\theta^{\langle t+1 \rangle}_{\langle t \rangle,1} - \theta^\star_{\langle t \rangle,1}| \leq (16\|\boldsymbol{\theta}^\star\| + 6)\|\boldsymbol{a}^{\langle t \rangle}\| + \kappa'_\theta | \|\boldsymbol{\theta}^{\langle t \rangle}\| - \theta^\star_{\langle t \rangle,1}|.
$$

So far we have proved in Equation (2.58) and Equation (2.68) and the following bounds:

$$
\begin{aligned}
|\theta^{\langle t+1 \rangle}_{\langle t \rangle,1} - \theta^\star_{\langle t \rangle,1}| &\leq (16\|\boldsymbol{\theta}^\star\| + 6)\|\boldsymbol{a}^{\langle t \rangle}\| + \kappa'_\theta | \|\boldsymbol{\theta}^{\langle t \rangle}\| - \theta^\star_{\langle t \rangle,1}| \\
|\theta^{\langle t+1 \rangle}_{\langle t \rangle,2} - \theta^\star_{\langle t \rangle,2}| &\leq \kappa_s |\theta^\star_{\langle t \rangle,2}|.
\end{aligned}
$$

Let $\kappa_\theta = \max\{\kappa_s, \kappa'_\theta\} \in (0,1)$ and $c'_\theta = 16\|\boldsymbol{\theta}^\star\| + 6$. Then, we conclude that

$$\begin{aligned}
\|\boldsymbol{\theta}^{\langle t+1\rangle} - \boldsymbol{\theta}^\star\|^2 &= |\theta^{\langle t+1\rangle}_{\langle t\rangle,1} - \theta^\star_{\langle t\rangle,1}|^2 + |\theta^{\langle t+1\rangle}_{\langle t\rangle,2} - \theta^\star_{\langle t\rangle,2}|^2 \\
&\leq ((16\|\boldsymbol{\theta}^\star\| + 6)\|\boldsymbol{a}^{\langle t\rangle}\| + \kappa'_\theta|\|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta^\star_{\langle t\rangle,1}|)^2 + (\kappa_s\theta^\star_{\langle t\rangle,2})^2 \\
&\leq \kappa_\theta^2(|\|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta^\star_{\langle t\rangle,1}|^2 + |\theta^\star_{\langle t\rangle,2}|^2) + (c'_\theta)^2\|\boldsymbol{a}^{\langle t\rangle}\|^2 + 2c'_\theta\|\boldsymbol{a}^{\langle t\rangle}\|\kappa_\theta|\|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta^\star_{\langle t\rangle,1}| \\
&\leq \kappa_\theta^2\|\boldsymbol{\theta}^{\langle t\rangle} - \boldsymbol{\theta}^\star\|^2 + ((c'_\theta)^2 + 2c'_\theta U_\theta + 2c'_\theta\|\boldsymbol{\theta}^\star\|)\|\boldsymbol{a}^{\langle t\rangle}\|.
\end{aligned}$$

Setting $c_\theta = (c'_\theta)^2 + 2c'_\theta U_\theta + 2c'_\theta\|\boldsymbol{\theta}^\star\|$ completes the proof of Lemma 2.12. □

2.4.4 Proof of Theorem 2.4 when $d \leq 2$

Due to Lemma 2.5, we can safely assume without loss of generality that $\langle\boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^\star\rangle > 0$ and $w^{\langle t\rangle} > 0.5$ for all $t > 0$. Then we use the following the strategy to prove Theorem 2.4.

1. Prove global convergence when the mean parameters $\boldsymbol{\theta}^\star$ is in one dimension.
2. Show that we can reduce the multi-dimensional problem into the one dimensional one.
3. Show geometric convergence by proving an attraction basin around $(\boldsymbol{\theta}^\star, w_1^\star)$.

2.4.4.1 One dimension case

In one dimension, the Population EM iterates of Model 4 follow the following update rule:

$$\begin{aligned}
w^{\langle t+1\rangle} &= G_w(\theta^{\langle t\rangle}, w_1^{\langle t\rangle}; \theta^\star, w_1^\star), \\
\theta^{\langle t+1\rangle} &= G_p(\theta^{\langle t\rangle}, w_1^{\langle t\rangle}; \theta^\star, w_1^\star).
\end{aligned} \tag{2.68}$$

Note that, we have a dynamic system of two coevolving sequences like Model 2. However, unlike Model 2, neither Model 1 nor Model 2 is the limit which Model 4 converges to, and therefore, we require a more general approach to analyze Model 4. Indeed, let us consider a general dynamic system defined as follows:

$$\begin{aligned} w^{\langle t+1\rangle} &= g_w(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \\ \theta^{\langle t+1\rangle} &= g_\theta(\theta^{\langle t\rangle}, w^{\langle t\rangle}), \end{aligned}$$

where $g_w(\theta, w)$ and $g_\theta(\theta, w)$ are two continuous functions. Our goal is to verify whether $\{(\theta^{\langle t\rangle}, w^{\langle t\rangle})\}$ converges to the fixed point $(\theta_\star, w_\star)$. Towards this goal, we establish the following conditions:

C.1 There exists a set $\mathcal{S} = (a_\theta, b_\theta) \times (a_w, b_w) \in \mathbb{R}^2$, where $a_\theta, b_\theta \in \mathbb{R} \cup \{\pm\infty\}$ and $a_w, b_w \in \mathbb{R}$, such that $\mathcal{S}$ contains point $(\theta_\star, w_\star)$ and point $(g_\theta(\theta, w), g_w(\theta, w)) \in \mathcal{S}$ for all $(\theta, w) \in \mathcal{S}$. Further, $g_\theta(\theta, w_1)$ is a non-decreasing function of θ for a given $w_1 \in (a_w, b_w)$ and $g_w(\theta, w_1)$ is a non-decreasing function of w for a given $\theta \in (a_\theta, b_\theta)$,

C.2 There is a *reference curve* $r\colon [a_w, b_w] \to [a_\theta, b_\theta]$ defined on $\bar{\mathcal{S}}$ (the closure of $\mathcal{S}$) such that:

C.2a r is continuous, decreasing, and passes through point $(\theta_\star, w_\star)$, i.e., $r(w_\star) = \theta_\star$.

C.2b Given $\theta \in (a_\theta, b_\theta)$, function $w \mapsto g_w(\theta, w)$ has a stable fixed point in $[a_w, b_w]$. Further, any stable fixed point w_s in $[a_w, b_w]$ or fixed point w_s in (a_w, b_w) satisfies the following:

(i) If $\theta < \theta_\star$ and $\theta \geq r(b_w)$, then $r^{-1}(\theta) > w_s > w_\star$.

(ii) If $\theta = \theta_\star$, then $r^{-1}(\theta) = w_s = w_\star$.

(iii) If $\theta > \theta_\star$ and $\theta \leq r(a_w)$, then $r^{-1}(\theta) < w_s < w_\star$.

C.2c Given $w \in [a_w, b_w]$, function $\theta \mapsto g_\theta(\theta, w)$ has a stable fixed point in $[a_\theta, b_\theta]$. Further, any stable fixed point θ_s in $[a_\theta, b_\theta]$ or fixed point θ_s in (a_θ, b_θ) satisfies the following:

(i) If $w_1 < w_\star$, then $r(w) > \theta_s > \theta_\star$.

(ii) If $w_1 = w_\star$, then $r(w) = \theta_s = \theta_\star$.

(iii) If $w_1 > w_\star$, then $r(w) < \theta_s < \theta_\star$.

We explain C.1 and C.2 in the right panel of Figure 2.2. Heuristically, we expect $(\theta^\star, w_1^\star)$ to be the only fixed point of the mapping $(\theta, w) \mapsto (g_\theta(\theta, w), g_w(\theta, w))$, and that $(\theta^{\langle t \rangle}, w^{\langle t \rangle})$ move toward this fixed point. Hence, we can prove the convergence of the iterates by showing certain geometric relationships between the curves of fixed points of the two functions. Hence, C.1 helps us to bound the iterates on the area that such nice geometric relations exist, and the reference curve r and C.2 are the tools to help us mathematically characterizing the geometric relations shown in the figure. Indeed, the next lemma implies that $C.1$ and $C.2$ are sufficient to show the convergence to the right point $(\theta_\star, w_\star)$:

Lemma 2.14. Suppose continuous functions $g_\theta(\theta, w), g_w(\theta, w)$ satisfy $C.1$ and $C.2$, then there exists a continuous mapping $m : \bar{\mathcal{S}} \to [0, \infty)$ such that $(\theta_\star, w_\star)$ is the only solution for $m(\theta, w) = 0$ on $\bar{\mathcal{S}}$, the closure of $\mathcal{S}$. Further, if we initialize $(\theta^{\langle 0 \rangle}, w^{\langle 0 \rangle})$ in $\mathcal{S}$, the sequence $\{(\theta^{\langle t \rangle}, w^{\langle t \rangle})\}_{t \geq 0}$ defined by

$$\theta^{\langle t+1 \rangle} = g_\theta(\theta^{\langle t \rangle}, w^{\langle t \rangle}), \quad \text{and} \quad w^{\langle t+1 \rangle} = g_w(\theta^{\langle t \rangle}, w^{\langle t \rangle}),$$

satisfies that $m(\theta^{\langle t \rangle}, w^{\langle t \rangle}) \downarrow 0$, and therefore $(\theta^{\langle t \rangle}, w^{\langle t \rangle})$ converges to $(\theta_\star, w_\star)$.

Proof. Based on $(\theta_\star, w_\star)$, we divide the region of $\mathcal{S} - \{(\theta_\star, w_\star)\}$ into 8 pieces:

- $R_1 = \{(\theta, w) \in \mathcal{S} : \theta \in [\theta_\star, \min\{r(a_w), b_\theta\}), w \in (a_w, w_\star]\} - \{(\theta_\star, w_\star)\}$.
- $R_2 = \{(\theta, w) \in \mathcal{S} : \theta \in [\theta_\star, \min\{r(a_w), b_\theta\}), w \in [w_\star, b_w)\} - \{(\theta_\star, w_\star)\}$.
- $R_3 = \{(\theta, w) \in \mathcal{S} : \theta \in (\max\{r(b_w), a_\theta\}, \theta_\star], w \in (a_w, w_\star]\} - \{(\theta_\star, w_\star)\}$.
- $R_4 = \{(\theta, w) \in \mathcal{S} : \theta \in (\max\{r(b_w), a_\theta\}, \theta_\star], w \in [w_\star, b_w)\} - \{(\theta_\star, w_\star)\}$.
- $R_5 = \{(\theta, w) \in \mathcal{S} : \theta \leq r(b_w), w \in (a_w, w_\star]\}$.
- $R_6 = \{(\theta, w) \in \mathcal{S} : \theta \leq r(b_w), w \in [w_\star, b_w)\}$.
- $R_7 = \{(\theta, w) \in \mathcal{S} : \theta \geq r(a_w), w \in (a_w, w_\star]\}$.
- $R_8 = \{(\theta, w) \in \mathcal{S} : \theta \geq r(a_w), w \in [w_\star, b_w)\}$.

Note that region R_5 to R_8 may not exists depending on the range of $r(w)$. Next, due to C.2a, we know the reference curve only crosses region R_1 and R_4. Note that $r^{-1}(\theta)$ exists on the regions R_1, R_2, R_3 and R_4. Hence, based on the points are above or below the reference curve r, we can further divide the region R_1 and R_4 into 4 pieces:

- $R_{11} = \{(\theta, w) \in R_1 : r^{-1}(\theta) \leq w\}$.
- $R_{12} = \{(\theta, w) \in R_1 : r^{-1}(\theta) \geq w\}$.
- $R_{41} = \{(\theta, w) \in R_4 : w \leq r^{-1}(\theta)\}$.
- $R_{42} = \{(\theta, w) \in R_4 : w \geq r^{-1}(\theta)\}$.

Now let's define $m : \mathcal{S} \to [0, \infty)$ based on the following 10 regions

$$\{R_{11}, R_{12}, R_2, R_3, R_{41}, R_{42}, R_5, R_6, R_7, R_8\} :$$

- If $(\theta, w) \in R_{11}$, $m(\theta, w) = (w_\star - w)(r(w) - \theta_\star)$, which is the area of the rectangle $D(\theta, w)$ given by $(\theta_\star, w_\star), (r(w), w)$.

- If $(\theta, w) \in R_{12}$, $m(\theta, w) = (w_\star - r^{-1}(\theta))(\theta - \theta_\star)$, which is the area of the rectangle $D(\theta, w)$ given by $(\theta_\star, w_\star), (\theta, r^{-1}(\theta))$.

- If $(\theta, w) \in R_2$, $m(\theta, w) = (w - r^{-1}(\theta))(\theta - r(w))$, which is the area of the rectangle $D(\theta, w)$ given by $(r(w), r^{-1}(\theta)), (\theta, w)$.

- If $(\theta, w) \in R_3$, $m(\theta, w) = (r^{-1}(\theta) - w)(r(w) - \theta)$, which is the area of the rectangle $D(\theta, w)$ given by $(r(w), r^{-1}(\theta)), (\theta, w)$.

- If $(\theta, w) \in R_{41}$, $m(\theta, w) = (r^{-1}(\theta) - w_\star)(\theta_\star - \theta)$, which is the area of the rectangle $D(\theta, w)$ given by $(\theta_\star, w_\star), (\theta, r^{-1}(\theta))$.

- If $(\theta, w) \in R_{42}$, $m(\theta, w) = (w - w_\star)(\theta_\star - r(w))$, which is the area of the rectangle $D(\theta, w)$ given by $(\theta_\star, w_\star), (r(w), w)$.

- If $(\theta, w) \in R_5$, $m(\theta, w) = (b_w - w)(r(w) - \theta)$, which is the area of the rectangle $D(\theta, w)$ given by $(r(w), b_w), (\theta, w)$.

- If $(\theta, w) \in R_6$, $m(\theta, w) = (b_w - w_\star)(\theta_\star - \theta)$, which is the area of the rectangle $D(\theta, w)$ given by $(\theta, b_w), (\theta_\star, w_\star)$.

- If $(\theta, w) \in R_7$, $m(\theta, w) = (w_\star - a_w)(\theta - \theta_\star)$, which is the area of the rectangle $D(\theta, w)$ given by $(\theta_\star, w_\star), (\theta, a_w)$.

- If $(\theta, w) \in R_8$, $m(\theta, w) = (w - a_w)(\theta - r(w))$, which is the area of the rectangle $D(\theta, w)$ given by $(r(w), a_w), (\theta, w)$.

It is straightforward to show that function m is a continuous function by checking the boundary and continuity of the reference function r. Further, $(\theta_\star, w_\star)$ is indeed

the only solution for $m(\theta, w) = 0$. Moreover, our construction of the rectangle D makes sure that

$$\text{If } (\tilde{\theta}, \tilde{w}) \text{ is strictly inside } D(\theta, w), \quad \text{then } D(\tilde{\theta}, \tilde{w}) \subsetneq D(\theta, w). \tag{2.69}$$

Next, we shall discuss the movement of the iterates from point $(\theta^{\langle t \rangle}, w^{\langle t \rangle})$ to point $(\theta^{\langle t+1 \rangle}, w^{\langle t+1 \rangle})$. For a given $w^{\langle t \rangle} \in [a_w, b_w]$, consider all the fixed points $\mathcal{V}$ in $[a_\theta, b_\theta]$ for $g_\theta(\theta, w)$ with respect to θ. Then, for any $\theta^{\langle t \rangle} \in (a_\theta, b_\theta)$,

- If $\theta^{\langle t \rangle}$ is a fixed point, then $\theta^{\langle t+1 \rangle}$ will stay at this fixed point.
- If $\theta^{\langle t \rangle}$ is not a fixed point, then we can find an interval $[q_1, q_2]$ such that
  - $\theta^{\langle t \rangle} \in [q_1, q_2]$ and $q_1, q_2 \in \mathcal{V} \bigcup \{a_\theta, b_\theta\}$
  - There is no fixed points in (q_1, q_2)
  - At least one of q_1 or q_2 is either a stable fixed point or one of a_θ, b_θ.

  Note that, since $g_\theta(\theta, w)$ is a non-decreasing function of θ and $(\theta^{\langle t+1 \rangle}, w^{\langle t+1 \rangle}) \in \mathcal{S}$, we know $\theta^{\langle t+1 \rangle} = g_\theta(\theta^{\langle t \rangle}, w^{\langle t \rangle}) \in [q_1, q_2]$ as well. Hence, comparing to the previous iteration $\theta^{\langle t \rangle}$, $\theta^{\langle t+1 \rangle} = g_\theta(\theta^{\langle t \rangle}, w^{\langle t \rangle})$ should move towards a stable fixed point q_i or a_θ, b_θ. Further, if $\theta^{\langle t+1 \rangle}$ moves towards a_θ or b_θ, then a_θ or b_θ has to be a stable fixed point as well. In other words, suppose $\theta^{\langle t+1 \rangle}$ move towards a_θ and a_θ is not a stable fixed point. Then a_θ is not a fixed point as well and there exists a constant $c > 0$ such that $\lim_{\theta \to a_\theta} g_\theta(\theta, w^{\langle t \rangle}) \leq a_\theta - c$. Hence by choosing θ close enough to a_θ, we know $g_\theta(\theta, w) < a_\theta$ which contradicts C.1.

In summary, we know the movement from $\theta^{\langle t \rangle}$ to $\theta^{\langle t+1 \rangle}$ is either stay at a fixed point or move towards a stable fixed point. Now, by C.2b, C.2c and discussing which region

(θ, w) belongs to, we can prove

$$\begin{aligned} &\text{Point } (\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle}) \text{ is } \text{ strictly inside } D(\theta^{\langle t\rangle}, w^{\langle t\rangle}), \\ &m(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle}) < m(\theta^{\langle t\rangle}, w^{\langle t\rangle}). \end{aligned} \tag{2.70}$$

and

$$\text{If } (\theta^{\langle t\rangle}, w^{\langle t\rangle}) \in R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4, \text{ then } (\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle}) \in R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4. \tag{2.71}$$

Note that depending on the regions, there are total 10 cases. But for simplicity, we show the proof for two cases: R_{11} and R_6 and leave the rest of the cases to the readers. For the first example, if point $(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \in R_{11}$, then we know there exists a fixed point $\theta_s \in [\theta_\star, b_\theta]$ for g_θ and $w_s \in [a_w, w_\star]$ for g_w such that $\theta^{\langle t+1\rangle} = g_\theta(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ lies in between $\theta^{\langle t\rangle}$ and θ_s, and $w^{\langle t+1\rangle} = g_w(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ lies in between $w^{\langle t\rangle}$ and w_s. Hence $(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})$ can only stay in R_1 which proves Equation (2.71) for the case $(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \in R_{11}$. Further, we have

$$|g_\theta(\theta^{\langle t\rangle}, w^{\langle t\rangle}) - \theta_s| \leq |\theta^{\langle t\rangle} - \theta_s|, \tag{2.72}$$

$$|g_w(\theta^{\langle t\rangle}, w^{\langle t\rangle}) - w_s| \leq |w^{\langle t\rangle} - w_s|, \tag{2.73}$$

where equality Equation (2.72)/Equation (2.73) holds if and only if $\theta^{\langle t\rangle} = \theta_s / w^{\langle t\rangle} = w_s$. Hence, by C.2, we have

- If $\theta^{\langle t\rangle} = \theta_\star$, then $w^{\langle t\rangle} < w_\star$. Hence we have $\theta_s \in (\theta_\star, r(w^{\langle t\rangle}))$ and $w_s = w_\star$. and therefore, Equation (2.73) is strict inequality. Hence, $w^{\langle t\rangle} < w^{\langle t+1\rangle}$.

- If $\theta^{\langle t\rangle} > \theta_\star$, then $\max(\theta_s, \theta^{\langle t\rangle}) \leq r(w^{\langle t\rangle})$ and $w_s > r^{-1}(\theta^{\langle t\rangle}) \geq w^{\langle t\rangle}$, therefore,

$$\theta^{\langle t+1\rangle} \;=\; g_\theta(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \leq r(w^{\langle t\rangle}), \quad \text{and} \quad w^{\langle t\rangle} < g_w(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \;=\; w^{\langle t+1\rangle}. \tag{2.74}$$

Therefore point $(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})$ lies in the rectangle $D(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ no matter what. Further, due to monotonic property of function r, we have

$$r(w^{\langle t\rangle}) > r(g_w(\theta^{\langle t\rangle}, w^{\langle t\rangle})). \tag{2.75}$$

Hence, by Equation (2.74) and Equation (2.75), no matter what region R_{11} or R_{12} contains the point $(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})$, the rectangle $D(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})$ is strictly smaller than the rectangle $D(\theta^{\langle t\rangle}, w^{\langle t\rangle})$. Hence, we have Equation (2.70) holds for the case $(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \in R_{11}$. For the second example that if $(\theta, w) \in R_6$, then by C.2, we know there exists a fixed point $\theta_s \in (r(b_w), \theta_\star]$ for g_θ and $w_s \in [w_\star, b_w]$ for g_w such that $\theta^{\langle t+1\rangle} = g_\theta(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ lies in between $\theta^{\langle t\rangle}$ and θ_s; and $w^{\langle t+1\rangle} = g_w(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ lies in between $w^{\langle t\rangle}$ and w_s. Hence, point $(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})$ can only stay in the region R_6 or R_4. Further, we have

$$|g_\theta(\theta^{\langle t\rangle}, w^{\langle t\rangle}) - \theta_s| \;\leq\; |\theta^{\langle t\rangle} - \theta_s|,$$

where equality holds if and only if $\theta^{\langle t\rangle} = \theta_s$. Therefore, we have

$$\theta^{\langle t+1\rangle} \;=\; g_\theta(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \;>\; \theta^{\langle t\rangle},$$

and hence, no matter what region R_6 or R_4 contains the point $(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})$, the rectangle $D(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})$ is strictly smaller than the rectangle $D(\theta^{\langle t\rangle}, w^{\langle t\rangle})$. Similarly,

we can show Equation (2.70) holds for all other cases. Next, we claim that if point $(\theta^{\langle 0\rangle}, w^{\langle 0\rangle}) \in R_5 \bigcup R_6 \bigcup R_7 \bigcup R_8$, then within finite steps t_0, the estimate $(\theta^{\langle t_0\rangle}, w^{\langle t_0\rangle})$ should lie in the region $R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4$. Suppose point $(\theta^{\langle 0\rangle}, w^{\langle 0\rangle}) \in R_6$, $g_\theta(\theta, w)/\theta$ is continuous on $[\theta^{\langle 0\rangle}, r(b_w)] \times [w_\star, b_w]$. Further, due to Equation (2.70), we have

$$g_\theta(\theta, w)/\theta \;\; > \;\; 1, \quad \forall (\theta, w) \in [\theta^{\langle 0\rangle}, r(b_w)] \times [w_\star, b_w].$$

Therefore, there exists a constant $\rho > 1$ such that $g_\theta(\theta, w) \geq \rho\theta$ on $[\theta^{\langle 0\rangle}, r(b_w)] \times [w_\star, b_w]$. Hence, within finite steps, we have $(\theta^{\langle t_0\rangle}, w^{\langle t_0\rangle}) \in R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4$. Similarly we can show for $(\theta^{\langle 0\rangle}, w^{\langle 0\rangle}) \in R_5, R_7, R_8$ as well. Hence, by Equation (2.71), we just need to focus on $(\theta^{\langle 0\rangle}, w^{\langle 0\rangle}) \in R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4$. Now we use contradiction to prove that $m(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ converges to 0. Suppose $m(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ does not converge to 0, then by definition of m, we know there exists some constant $c_\theta > 0$ and $c_w > 0$, such that

$$|\theta_\star - \theta^{\langle t\rangle}| \geq c_\theta \quad \text{and} \quad |w_\star - w^{\langle t\rangle}| \geq c_w, \quad \forall t \geq 0. \tag{2.76}$$

Further, since $\mathcal{S} \supset D(\theta^{\langle 0\rangle}, w^{\langle 0\rangle}) \supset D(\theta^{\langle 1\rangle}, w^{\langle 1\rangle}) \supset \cdots$, we know all points $(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ are bounded on a compact set $D(\theta^{\langle 0\rangle}, w^{\langle 0\rangle})$. Now consider function

$$U(\theta^{\langle t\rangle}, w^{\langle t\rangle}) := \frac{m(\theta^{\langle t+1\rangle}, w^{\langle t+1\rangle})}{m(\theta^{\langle t\rangle}, w^{\langle t\rangle})}$$

we know U is continuous on $(\theta^{\langle t\rangle}, w^{\langle t\rangle}) \in Q = \{(\theta, w) \in D(\theta^{\langle 0\rangle}, w^{\langle 0\rangle}) : |\theta_\star - \theta| \geq c_\theta, |w_\star - w| \geq c_w\}$. Further, since Q is a compact set and $U < 1$ on Q, we know there exists constant $\rho < 1$ such that $\sup_Q U(\theta, w) \leq \rho$. Hence, we have $m(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ converges to 0. Therefore, $(\theta^{\langle t\rangle}, w^{\langle t\rangle})$ converges to $(\theta_\star, w_\star)$ since it is the only solution

for $m = 0$ and m is continuous. □

Remark 2.3. The contradiction proof above is an analog to the proof for Model 1 in Section 2.4.1, and again this strategy does not guarantee geometric convergence until we can analyze locally around the fixed point.

Remark 2.4. We consider the construction of function m is an extension to the proof of Model 1. Indeed, the key property Equation (2.34) implies the following statements: Suppose $\theta^{\langle t\rangle} > 0$, then $\theta^{\langle t+1\rangle}$ will be strictly inside the interval $[\theta^{\langle t\rangle}, \theta^\star]$ or $[\theta^\star, \theta^{\langle t\rangle}]$. Therefore, the length of the interval defined by $\theta^{\langle t\rangle}$ and $\theta^\star$ is strictly decreasing. Hence from the proof of Lemma 2.14, it is clear that the rectangle $D(\theta, w)$ is an extension of the interval $[\theta, \theta^\star]$ and the function m is an extension to the length of the interval. On the other hand, we here consider m as the areas of the rectangles. From the rectangles, one can also construct upper bounds of the ℓ_2 distance between the estimate to $(\theta_\star, w_\star)$ and show this sequence of upper bounds converge to 0.

In our case of Model 4, we let $g_\theta(\theta, w_1)$ and $g_w(\theta, w_1)$ be the shorthand for the two update functions $G_\theta(\theta, w_1; \theta^\star, w_1^\star)$ and $G_w(\theta, w_1; \theta^\star, w_1^\star)$ in Equation (2.68) for a fixed $(\theta^\star, w_1^\star)$. Also, we set $a_w = 0.5, b_w = 1, a_\theta = 0, b_\theta = \infty$ and $(\theta_\star, w_\star) = (\theta^\star, w_1^\star)$. Then we just need to verify C.1 and C.2.

To verify C.1, note that

$$\begin{aligned}
\frac{\partial g_w(\theta, w_1)}{\partial w_1} &= \int \frac{\phi(y-\theta)\phi(y+\theta)}{(w_1\phi(y-\theta) + w_2\phi(y+\theta))^2}\phi^+(y, \theta, w_1^\star)\mathrm{d}y \;>\; 0, \\
\frac{\partial g_\theta(\theta, w_1)}{\partial \theta} &= \int \frac{4w_1w_2}{(w_1\phi(y-\theta) + w_2\phi(y+\theta))^2}\phi^+(y, \theta, w_1^\star)\mathrm{d}y \;\geq\; 0.
\end{aligned}$$

Hence, g_w is a increasing function of w_1 for all given θ and g_θ is a non-decreasing function of θ for all given w_1. The rest of C.1 is guaranteed by our assumption stated in the beginning (See Lemma 2.5).

To show C.2, we first define the reference curve r by

$$r(w_1) := \frac{w_1^\star - w_2^\star}{w_1 - w_2}\theta^\star = \frac{2w_1^\star - 1}{2w_1 - 1}\theta^\star, \qquad \forall w_1 \in (0.5, 1], w_2 = 1 - w_1. \tag{2.77}$$

The claim C.2a holds by construction. To show C.2b, according to Equation (2.77), function r is a one to one mapping between $w \in (0.5, 1]$ and $\theta \in [(w_1^\star - w_2^\star)\theta^\star, \infty)$. Hence, we can simplify C.2b as

- If $w_1 \in (w_1^\star, 1]$, then $w_1 > w_s > w_1^\star$,
- If $w_1 = w_1^\star$, then $w_1 = w_s = w_\star$,
- If $w_1 \in (0.5, w_1^\star)$, then $w_1 < w_s < w_1^\star$,

where w_s is any stable fixed point in $[a_w, b_w]$ or fixed point in (a_w, b_w) for $\theta = r(w_1)$. To prove the claim, we establish an even stronger property of the weights update function $g_w(\theta, w)$: for any fixed $\theta > 0$, the function $w_1 \mapsto g_w(\theta, w_1)$ has at most one other fixed point besides $w_1 = 0$ and $w_1 = 1$, and most importantly, it has only one unique stable fixed point. This is formalized in the following lemma.

Lemma 2.15. For all $\theta > 0$, there are at most three fixed points for $g_w(\theta, w_1)$ with respect to w_1. Further, there exists an unique stable fixed point $F_w(\theta) \in (0, 1]$, i.e., (i) $F_w(\theta) = g_w(\theta, F_w(\theta))$ and (ii) for all $w_1 \in (0, 1)$, we have

$$g_w(\theta, w_1) < w_1 \Leftrightarrow w_1 < F_w(\theta) \quad \text{and} \quad g_w(\theta, w_1) > w_1 \Leftrightarrow w_1 > F_w(\theta). \tag{2.78}$$

We prove this lemma in Appendix A.1.4.1.

We explain Lemma 2.15 in Figure 2.1. Note that, in the figure, we observe that g_w is an increasing function with $g_w(\theta, 0) = 0$ and $g_w(\theta, 1) = 1$. Further, it is ei-

ther a concave function, it is piecewise concave-then-convex[2]. Hence, we know if $\partial g_w(\theta, w_1)/\partial w_1\big|_{w_1=1}$ is at most 1, the only stable fixed point is $w_1 = 1$, else if the derivative is larger than 1, there exists only one fixed point in (0,1) and it is the only stable fixed point.

By Equation (2.78) in Lemma 2.15, we can complete the proof for C.2b by showing the following technical lemma

Lemma 2.16. Let $\gamma = \frac{2w_1^\star - 1}{2w_1 - 1}$, we have

$$\begin{aligned} g_w(\gamma\theta^\star, w_1) &< w_1 \quad \text{and} \quad g_w(\gamma\theta^\star, w_1^\star) > w_1^\star \quad \forall w_1 \in (w_1^\star, 1] \\ g_w(\gamma\theta^\star, w_1) &> w_1 \quad \text{and} \quad g_w(\gamma\theta^\star, w_1^\star) < w_1^\star \quad \forall w_1 \in (0.5, w_1^\star) \end{aligned}$$

We prove this lemma in Appendix A.1.4.2.

The final step to apply Lemma 2.14 is to prove C.2c. However, $(\theta, w_1) = ((2w_1^\star - 1)\theta^\star, 1)$ is a point on the reference curve r and $\theta = (2w_1^\star - 1)\theta^\star$ is a stable fixed point for $g_\theta(\theta, 1)$. This violates C.2c. To address this issue, since we can characterize the shape and the number of fixed points for g_w, by typical uniform continuity arguments, we can find $\delta, \epsilon > 0$ such that the adjusted reference curve $r_{adj}(w) := r(w) - \epsilon \cdot \max(0, w - 1 + \delta)$ satisfies C.2a and C.2b. Then we can prove that the adjusted reference curve $r_{adj}(w)$ satisfies C.2c. Specifically, note that, we have

$$r_{adj}(w) \;=\; r(w) - \epsilon \cdot \max(0, w - 1 + \delta) \;=\; \frac{2w_1^\star - 1}{2w - 1}\theta^\star - \epsilon \cdot \max(0, w - 1 + \delta),$$

for some positive $\epsilon, \delta > 0$. Also, note that $g_\theta(\theta, 1) \equiv (2w_1^\star - 1)\theta^\star$. Hence, we just need to show the following

C.2c' Given $w_1 \in (a_w, b_w)$, any stable fixed point θ_s of $g_\theta(\theta, w)$ in $[a_\theta, b_\theta]$ or fixed point

[2]There exists $\tilde{w} \in (0, 1)$ such that $g_w(\theta, w)$ is concave in $[0, \tilde{w}]$ and convex in $[\tilde{w}, 1]$.

θ_s in (a_θ, b_θ) satisfies that

- If $w_1 < w_\star$, then $r(w) > \theta_s > \theta_\star$.
- If $w_1 = w_\star$, then $r(w) = \theta_s = \theta_\star$.
- If $w_1 > w_\star$, then $r(w) < \theta_s < \theta_\star$.

We first show that there exists stable fixed point for $g_\theta(\theta, w_1)$ with respect to θ, i.e.,

Claim 1 If $w_1 \in (0.5, w_1^\star]$, then there exists an unique non-negative fixed point for $g_\theta(\theta, w_1)$ denoted as $F_\theta(w_1)$. Further, $F_\theta(w_1) \geq \theta^\star$.

Claim 2 If $w_1 \in (w_1^\star, 1]$, then there exists positive stable fixed point for $g_\theta(\theta, w_1)$ and all non-negative fixed points are in $(0, \theta^\star)$.

First, it is clear that $\theta = 0$ is not a fixed point for $w_1 > 0.5$ and $w_1^\star > 0.5$, therefore, we just need to consider $\theta > 0$. Then, to prove Claim 1 and Claim 2, we should find out the shape of $g_\theta(\theta, w_1)$ for different true values $(\theta^\star, w_1^\star)$. Notice that, by Lemma 2.2, we know the shape of $H(\theta; \theta^\star, w_1) = G_\theta(\theta, w_1; \theta^\star, w_1)$, i.e., for $\theta > 0, w_1 \in [0.5, 1]$

$$H(\theta; \theta^\star, w_1) \quad \gtreqless \quad \theta \quad \text{is equivalent to} \quad \theta \lesseqgtr \theta^\star. \tag{2.79}$$

Hence, our next step to compare $G_\theta(\theta, w_1; \theta^\star, w_1^\star)$ with $H(\theta; \theta^\star, w_1) = G_\theta(\theta, w_1; \theta^\star, w_1)$. Note that, we have

$$\begin{aligned}
\frac{\partial G_\theta(\theta, w_1; \theta^\star, w_1^\star)}{\partial w_1^\star} &= \int y \frac{w_1 e^{y\theta} - w_2 e^{-y\theta}}{w_1 e^{y\theta} + w_2 e^{-y\theta}} \left(\phi(y - \theta^\star) - \phi(y + \theta^\star)\right) \mathrm{d}y \\
&= \int_{y \geq 0} \left(\frac{w_1 e^{y\theta} - w_2 e^{-y\theta}}{w_1 e^{y\theta} + w_2 e^{-y\theta}} + \frac{w_1 e^{-y\theta} - w_2 e^{y\theta}}{w_1 e^{-y\theta} + w_2 e^{y\theta}} \right) y \left(\phi(y - \theta^\star) - \phi(y + \theta^\star)\right) \mathrm{d}y \\
&= 2 \int_{y \geq 0} \frac{w_1 - w_2}{\left(w_1 e^{y\theta} + w_2 e^{-y\theta}\right)\left(w_1 e^{-y\theta} + w_2 e^{y\theta}\right)} y \left(\phi(y - \theta^\star) - \phi(y + \theta^\star)\right) \mathrm{d}y \;>\; 0.
\end{aligned} \tag{2.80}$$

Hence, if $w_1 \in (w_1^\star, 1]$, we know G_θ will be strictly below H. Therefore

$$G_\theta(\theta, w_1; \theta^\star, w_1^\star) \ < \ \theta, \quad \forall \theta \geq \theta^\star.$$

Hence, with $G_\theta(0, w_1; \theta^\star, w_1^\star) = (w_1 - w_2)(w_1^\star - w_2^\star)\theta^\star > 0$ and continuity of the function, we know Claim 2 holds. Similarly, if $w_1 \in (0.5, w_1^\star]$, we know G_θ will be strictly above H. Therefore

$$G_\theta(\theta, w_1; \theta^\star, w_1^\star) \ > \ \theta, \quad \forall 0 < \theta \leq \theta^\star.$$

Hence, to prove Claim 1, we just need to show that $G_\theta(\theta, w_1; \theta^\star, w_1^\star)$ is bounded by some constant C and

$$\frac{\partial G_\theta(\theta, w_1; \theta^\star, w_1^\star)}{\partial \theta} \ < \ 1, \quad \forall \theta \geq \theta^\star, w_1 \in (0.5, w_1^\star]. \tag{2.81}$$

To prove boundedness, we have the following more general lemma:

Lemma 2.17. Given any $(\boldsymbol{\theta}, w_1, \boldsymbol{\theta}^\star, w_1^\star)$, we have

$$\|G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star)\|^2 \ \leq \ 1 + \|\boldsymbol{\theta}^\star\|^2.$$

Hence, for all $t \geq 1$, $\|\boldsymbol{\theta}^{\langle t \rangle}\|^2 \leq \|\boldsymbol{\theta}^\star\|^2 + 1$.

We prove this lemma in Appendix A.1.4.3.

To prove Equation (2.81), we have for $\theta \geq \theta^\star$,

$$\begin{aligned}
\frac{\partial G_\theta(\theta, w_1; \theta^\star, w_1^\star)}{\partial \theta} &= \int \frac{4w_1 w_2}{\left(w_1 e^{y\theta} + w_2 e^{-y\theta}\right)^2} y^2 \left(w_1^\star \phi(y - \theta^\star) + w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y \\
&= \frac{\partial H(\theta; \theta^\star, w_1)}{\partial \theta} + (w_1^\star - w_1) \int \frac{4w_1 w_2}{\left(w_1 e^{y\theta} + w_2 e^{-y\theta}\right)^2} y^2 \left(\phi(y - \theta^\star) - \phi(y + \theta^\star)\right) \mathrm{d}y \\
&\overset{(i)}{\leq} \frac{\partial H(\theta; \theta^\star, w_1)}{\partial \theta} \\
&\overset{(ii)}{\leq} e^{-\frac{(\theta^\star)^2}{2}} < 1,
\end{aligned}$$

where inequality (ii) holds due to Lemma 2.2 and inequality (i) holds due to

$$w_1 e^{y\theta} + w_2 e^{-y\theta} \geq w_1 e^{-y\theta} + w_2 e^{y\theta}, \quad \forall \theta > 0.$$

This completes the proof for Claim 1 and Claim 2. Finally, it is straightforward to show the rest of C.2c by Claim 1 and Claim 2 and the following lemma:

Lemma 2.18.

$$g_\theta(\gamma\theta^\star, w_1) < \gamma\theta^\star, \quad \forall w_1 \in (\frac{1}{2}, w_1) \tag{2.82}$$

$$g_\theta(b\theta^\star, w_1) > b\theta^\star, \quad \forall b \in (0, \gamma], w_1 \in (w_1, 1). \tag{2.83}$$

We prove this lemma in Appendix A.1.4.4.

Remark 2.5. We should point out that despite of the illustration from the right panel of Figure 2.2, we have not proved that the fixed points of g_θ forms a continuous curve as a function of w_1 and the fixed points of g_w forms a continuous curve as a function of θ. One advantage of Lemma 2.14 is its capability to handle the case when continuous fixed point curves may not exist. On the other hand, when two fixed point curves both exist, we can show convergence directly without the construction of the reference

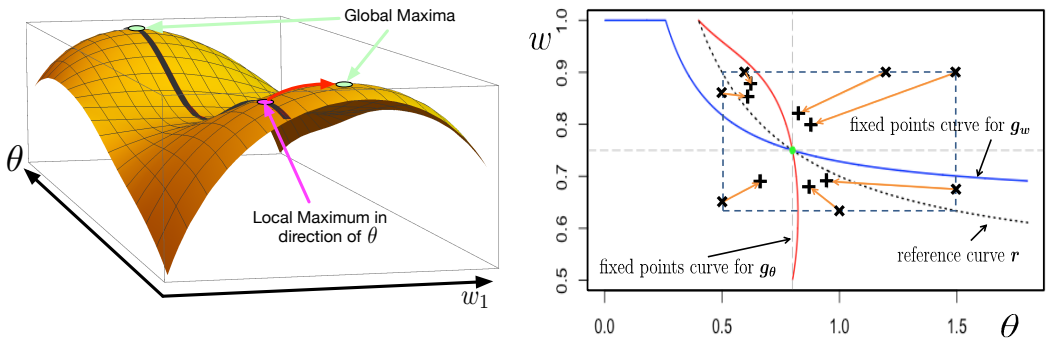

Figure 2.2: Left panel: The landscapes of log-likelihood objectives for Population EM_1 and Population EM_2 with $(\boldsymbol{\theta}^\star, w_1^\star) = (1, 0.4)$ are shown in the black belt and the yellow surface respectively. The two green points indicates the two global maxima of Population EM_2, one of which is also the global maximum of Population EM_1. The purple point indicates the local maximum of Population EM_1. Over-parameterization helps us to escape the local maximum through the direction of w_1. Right panel: The fixed point curves for functions g_θ and g_w are shown with red and blue lines respectively. The green point at the intersections of the three curves is the correct convergence point $(\theta_\star, w_\star)$. The black dotted curve shows the reference curve r. The cross points $\times$ are the possible initializations and the plus points $+$ are the corresponding positions after the first iteration. By the geometric relations between the three curves, the iterations have to converge to $(\theta_\star, w_\star)$

curve. Indeed, we will discuss it further in Chapter 3.

2.4.4.2 Reduction to one dimension case

In this section, we show how to reduce multi-dimensional problem into one-dimensional problem by proving the angle between the two vectors $\boldsymbol{\theta}^\star$ and $\boldsymbol{\theta}^{\langle t \rangle}$ is decreasing to 0. We use similar strategy shown in Section 2.4.3 to prove this. Specifically, let us recall the definition of $\alpha^{\langle t \rangle}$ and $\beta^{\langle t \rangle}$ in Section 2.4.3, i.e., the angle between the two vectors $\boldsymbol{\theta}^{\langle t \rangle}$ and $\boldsymbol{\theta}^{\langle t+1 \rangle}$ and the angle between the two vectors $\boldsymbol{\theta}^{\langle t \rangle}$ and $\boldsymbol{\theta}^\star$ respectively. Then we will first show that $|\alpha^{\langle t \rangle}| \leq |\beta^{\langle t \rangle}|$ and $\{|\beta^{\langle t \rangle}|\}$ is a non-increasing sequence and then show that $|\beta^{\langle t \rangle}| \to 0$ as $t \to \infty$. Towards this goal, since Lemma 2.3 holds for Model 4

as well, just like the analysis for Model 2, we are free to apply any coordinate system to prove any statements that are invariant under rotation. Hence, let us recall the sequence of the coordinate systems $\mathcal{A}$ where at iteration t, the coordinates are chosen such that $\boldsymbol{\theta}^{\langle t \rangle} = (\|\boldsymbol{\theta}^{\langle t \rangle}\|, 0)^\top$. Further, we adopt the notation $\boldsymbol{\theta}^{\langle t+1 \rangle}_{\langle t \rangle} = (\theta^{\langle t+1 \rangle}_{\langle t \rangle,1}, \theta^{\langle t+1 \rangle}_{\langle t \rangle,2})$ for $\boldsymbol{\theta}^{\langle t+1 \rangle}$ under the coordinates with respect to $\boldsymbol{\theta}^{\langle t \rangle}$. Finally, due to Lemma 2.3, $\{w^{\langle t \rangle}\}$ remains the same under rotation, we keep the same notation for $w^{\langle t \rangle}$ when we apply the coordinate systems $\mathcal{A}$.

Then given $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle > 0$, we have

- If $\beta^{\langle 0 \rangle} = 0$, then for $t \geq 1$, we have $\beta^{\langle t \rangle} = 0$, i.e., it is an one-dimensional problem.
- If $|\beta^{\langle 0 \rangle}| \in (0, \frac{\pi}{2})$, then for $t \geq 1$, we have $|\beta^{\langle t \rangle}| \in (0, |\beta^{\langle t-1 \rangle}|)$.

We assume $\beta^{\langle 0 \rangle} \geq 0$ and the proof for $\beta^{\langle 0 \rangle} < 0$ is similar. Further, it is straightforward to verify that if $\beta^{\langle 0 \rangle} = 0$, we have $\beta^{\langle t \rangle} = 0, \forall t \geq 0$. Therefore, we just need to show $\beta^{\langle t \rangle} < \beta^{\langle t-1 \rangle}, \forall t > 0$. To prove this, we just need to to prove the same three statements as in Section 2.4.3 hold for $\forall t \geq 0$:

(i) $\beta^{\langle t \rangle} \in (0, \frac{\pi}{2})$.

(ii) $\alpha^{\langle t \rangle} \in (0, \beta^{\langle t \rangle})$.

(iii) $\beta^{\langle t+1 \rangle} = \beta^{\langle t \rangle} - \alpha^{\langle t \rangle} \in (0, \beta^{\langle t \rangle})$.

We use induction to show (i)-(iii) by proving the following chain of arguments:

Claim 1 If (i) holds for t, then (ii) holds for t.

Claim 2 If (i) and (ii) hold for t, then (iii) holds for t.

Claim 3 If (i), (ii), and (iii) hold for t, then (i) holds for $t + 1$.

(i) holds for $t = 0$ and Claim 2 and 3 are trivially true. So we just have to prove Claim 1 for all $t \geq 0$. It is straightforward to check that the Claims are invariant under any rotation of the coordinates. Hence, due to Lemma 2.3, we use the sequence of the coordinate systems $\mathcal{A}$. Under $\mathcal{A}$, we apply the same calculations in Equation (A.20), Equation (A.22) and Equation (A.23)(where $(\tilde{\theta}_1, \tilde{\theta}_2)$ and $(\theta^\star_\parallel, \theta^\star_\perp)$ correspond to $(\theta^{\langle t+1\rangle}_{\langle t\rangle,1}, \theta^{\langle t+1\rangle}_{\langle t\rangle,2})$ and $(\theta^\star_{\langle t\rangle,1}, \theta^\star_{\langle t\rangle,2})$ respectively), and obtain that

$$0 \;<\; \tan\alpha^{\langle t\rangle} \;<\; \tan\beta^{\langle t\rangle} \;=\; \theta^\star_{\langle t\rangle,2}/\theta^\star_{\langle t\rangle,1}.$$

Hence, we have Claim 1 holds.

Next, we want to prove the angle $|\beta^{\langle t\rangle}|$ is decreasing to 0. Again, we assume $\beta^{\langle 0\rangle} \geq 0$ and continue to apply the sequence of the coordinate systems $\mathcal{A}$. To show $\beta^{\langle t\rangle}$ decreases to 0, it is equivalent to show that $\theta^\star_{\langle t\rangle,1}$ converges to $\|\boldsymbol{\theta}^\star\|$. Then, since $\beta^{\langle t\rangle}$ is decreasing, we have $\theta^\star_{\langle t\rangle,1} = \|\boldsymbol{\theta}^\star\| \cdot \beta^{\langle t\rangle}$ is increasing. Hence

$$\theta^\star_{\langle t\rangle,1} \in [\theta^\star_{\langle 1\rangle,1}, \|\boldsymbol{\theta}^\star\|], \quad \forall t \geq 1. \tag{2.84}$$

To prove the increasing sequence $\theta^\star_{\langle t+1\rangle,1}$ converges to $\|\boldsymbol{\theta}^\star\|$, we just need to show that for any $\hat{\theta} < \|\boldsymbol{\theta}^\star\|$, we can find $\theta^\star_{\langle t+1\rangle,1}/\theta^\star_{\langle t\rangle,1} \geq \rho_{\hat{\theta}}$ for some constant $\rho_{\hat{\theta}} > 1$, then with a straightforward contradiction argument, within finite iterations, we should have $\theta^\star_{\langle t'\rangle,1} > \hat{\theta}$ for a certain t', which implies $\theta^\star_{\langle t+1\rangle,1}$ converges to $\|\boldsymbol{\theta}^\star\|$. To find such ρ, note that, since $\theta^\star_{\langle t\rangle,1}$ is a value invariant to coordinate rotations, by Equation (A.20),Equation (A.22) and Equation (A.23), we have $\theta^\star_{\langle t+1\rangle,1}/\theta^\star_{\langle t\rangle,1}$, as a function of $\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}$ and $\theta^\star_{\langle t\rangle,1}$, is continuous and

$$\theta^\star_{\langle t+1\rangle,1}/\theta^\star_{\langle t\rangle,1} \;>\; 1, \quad \forall \|\boldsymbol{\theta}^{\langle t\rangle}\| > 0, w_1^{\langle t\rangle} \in (0.5, 1], \theta^\star_{\langle t\rangle,1} \in [\theta^\star_{\langle 1\rangle,1}, \|\boldsymbol{\theta}^\star\|).$$

Hence, we just need to find some constants $0 < c_1 < c_2$ and $0.5 < c_3 < 1$ such that $\|\boldsymbol{\theta}^{\langle t \rangle}\| \in [c_1, c_2]$ and $w_1^{\langle t \rangle} \in [c_3, 1]$ for $t \geq 1$, then we can find ρ by the uniform continuity argument. From Lemma 2.17, we have $c_2 = 1 + \|\boldsymbol{\theta}^\star\|$. Since both $\|\boldsymbol{\theta}^{\langle t \rangle}\|$ and $w_1^{\langle t \rangle}$ is invariant to the coordinate rotations, we continue to apply the sequence of coordinate systems $\mathcal{A}$. Note that, we have

$$\begin{aligned}
\theta_{\langle t \rangle,1}^{\langle t+1 \rangle} &= \int y \frac{w_1^{\langle t \rangle} e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} - w_2^{\langle t \rangle} e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}}{w_1^{\langle t \rangle} e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} + w_2^{\langle t \rangle} e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}} \left(w_1^\star \phi(y - \theta_{\langle t \rangle,1}^\star) + w_2^\star \phi(y + \theta_{\langle t \rangle,1}^\star) \right) \mathrm{d}y \\
&= G_\theta(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle}; \theta_{\langle t \rangle,1}^\star, w_1^\star) \\
w_1^{\langle t+1 \rangle} &= \int \frac{w_1^{\langle t \rangle} e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} - w_2^{\langle t \rangle} e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}}{w_1^{\langle t \rangle} e^{y\|\boldsymbol{\theta}^{\langle t \rangle}\|} + w_2^{\langle t \rangle} e^{-y\|\boldsymbol{\theta}^{\langle t \rangle}\|}} \left(w_1^\star \phi(y - \theta_{\langle t \rangle,1}^\star) + w_2^\star \phi(y + \theta_{\langle t \rangle,1}^\star) \right) \mathrm{d}y \\
&= G_w(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle}; \theta_{\langle t \rangle,1}^\star, w_1^\star) \qquad (2.85)
\end{aligned}$$

Hence, $(\theta_{\langle t \rangle,1}^{\langle t+1 \rangle}, w_1^{\langle t+1 \rangle})$ is the next iteration of $(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle})$ of the Population EM estimates under the true value $(\theta_{\langle t \rangle,1}^\star, w_1^\star)$. Indeed, we can consider this two dimensional problem as a series of one dimensional problems that follows this procedure:

Step 1 Start with point $(\|\boldsymbol{\theta}^{\langle 1 \rangle}\|, w_1^{\langle 1 \rangle}) \in S$, where $S = (0, \infty) \times (0.5, 1)$.

Step 2 For iteration t, let point $(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle})$ move towards the point $(\theta_{\langle t \rangle,1}^{\langle t+1 \rangle}, w_1^{\langle t+1 \rangle})$ following the one dimensional update rule for the true value $\theta_\star = \theta_{\langle t \rangle,1}^\star$.

Step 3 Shift the true value $\theta_\star = \theta_{\langle t \rangle,1}^\star$ and the point $(\theta_{\langle t \rangle,1}^{\langle t+1 \rangle}, w_1^{\langle t+1 \rangle})$ to the right to their new values: true value $\theta_\star = \theta_{\langle t+1 \rangle,1}^\star$ and new point $(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle})$.

Step 4 End iteration t and go back to Step 2 for iteration $t + 1$.

To analyze this, we recall our analysis for the one dimension case previously in this section. Due to Lemma 2.15 holds for any non-zero true value $\theta^\star$, by typical uniform continuity argument, we can find $\delta, \epsilon > 0$ such that the adjusted reference

curve $r_{adj}(w_1; \theta_\star)$ defined by

$$r_{adj}(w_1; \theta_\star) \;=\; \frac{2w_1^\star - 1}{2w_1 - 1}\theta_\star - \epsilon \cdot \max(0, w_1 + \delta - 1) > 0,$$

satisfies C.1,C.2 with $(a_\theta, b_\theta) = (0, \infty), (a_w, b_w) = (0.5, 1)$ for any true value $\theta_\star \in [\theta^\star_{\langle 1\rangle,1}, \|\boldsymbol{\theta}^\star\|]$ and $w_\star = w_1^\star$. Hence, on $S = (0, \infty) \times (0.5, 1)$, as $\theta_\star$ increases, the reference curve shifted to the right. Further, for any point (θ, w) in S, recall its corresponding area function $m(\theta, w)$ and rectangle $D(\theta, w)$ in the proof for Lemma 2.14. We use $m(\theta, w; \theta_\star)$ and $D(\theta, w; \theta_\star)$ to denote their values under the true value $\theta_\star$. By their definitions, we note that the left side and down side of the rectangle $D(\theta, w; \theta_\star)$ is non-decreasing as $\theta_\star$ increases. Hence, by Equation (2.70), we know as $\theta^\star_{\langle t\rangle,1}$ increases, $w_1^{\langle t\rangle}$ is always lower bounded by the down side of the rectangle $D(\|\boldsymbol{\theta}^{\langle 1\rangle}\|, w_1^{\langle 1\rangle}; \theta^\star_{\langle 1\rangle,1})$ due to the following chain of arguments:

$$\begin{aligned}
w_1^{\langle t+1\rangle} &\overset{\text{(i)}}{\geq} \text{ bottom side of } D(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta^\star_{\langle t\rangle,1}) \\
&\overset{\text{(ii)}}{\geq} \text{ bottom side of } D(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta^\star_{\langle t-1\rangle,1}) \\
&\overset{\text{(iii)}}{\geq} \text{ bottom side of } D(\theta^{\langle t\rangle}_{\langle t-1\rangle,1}, w_1^{\langle t-1\rangle}; \theta^\star_{\langle t-1\rangle,1}) \\
&\overset{\text{(iv)}}{\geq} \text{ bottom side of } D(\|\boldsymbol{\theta}^{\langle t-1\rangle}\|, w_1^{\langle t-1\rangle}; \theta^\star_{\langle t-1\rangle,1}) \\
&\geq \quad \cdots \;\geq\; \text{bottom side of } D(\|\boldsymbol{\theta}^{\langle 1\rangle}\|, w_1^{\langle 1\rangle}; \theta^\star_{\langle 1\rangle,1}) \;=\; c_3,
\end{aligned}$$

where inequality (i) holds due to Equation (2.70), inequality (ii) and (iii) hold due to the shift of reference curve and definition of the rectangle D, and inequality (iv) holds due to Equation (2.69). Also, we can show

$$\|\boldsymbol{\theta}^{\langle t\rangle}\| \geq \min\{\|\boldsymbol{\theta}^{\langle 1\rangle}\|, (w_1^\star - w_2^\star)\theta^\star_{\langle 1\rangle,1} - \epsilon\delta\} := c_1.$$

This is because,

- If $\|\boldsymbol{\theta}^{\langle t\rangle}\| \leq \theta^{\star}_{\langle t\rangle,1} - \epsilon\delta$, i.e., point $(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle})$ is inside the region R_5 or R_6 defined by the true value $\theta_\star = \theta^{\star}_{\langle t\rangle,1}$, then we know $\|\boldsymbol{\theta}^{\langle t+1\rangle}\| \geq \theta^{\langle t+1\rangle}_{\langle t\rangle,1} \geq \|\boldsymbol{\theta}^{\langle t\rangle}\|$.

- If $\|\boldsymbol{\theta}^{\langle t\rangle}\| \leq \theta^{\star}_{\langle t\rangle,1} - \epsilon\delta$, i.e., point $(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle})$ is inside the regions R_1-R_4 (note that regions R_7 and R_8 doesn't exists here), we have $(\boldsymbol{\theta}^{\langle t+1\rangle}_{\langle t\rangle}, w_1^{\langle t+1\rangle})$ stay at R_1-R_4 and hence $\|\boldsymbol{\theta}^{\langle t+1\rangle}\| \geq \theta^{\langle t+1\rangle}_{\langle t\rangle,1} \geq \theta^{\star}_{\langle t\rangle,1} - \epsilon\delta$.

Hence, this completes the proof of our claim that the angle $\beta^{\langle t\rangle}$ is decreasing to 0.

Finally, we want to show that $(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle})$ converges to $(\|\boldsymbol{\theta}^{\star}\|, w_1^{\star})$ which implies $(\boldsymbol{\theta}^{\langle t\rangle}, w_1^{\langle t\rangle})$ converges to $(\boldsymbol{\theta}^{\star}, w_1^{\star})$ due to $\beta^{\langle t\rangle} \to 0$. Since the ℓ_2 norm is rotation invariant, we apply the sequence of the coordinate system $\mathcal{A}$. To prove this final step, we just need to bound $w_1^{\langle t\rangle}$ away from 1, i.e., there exists $c_4 \in (0,1)$ such that

$$w_1^{\langle t\rangle} \;\leq\; c_4 \;<\; 1, \quad \forall t \geq 1. \tag{2.86}$$

Note that if Equation (2.86) holds. Consider the following functions

$$\begin{aligned}
U_1 &= m(\theta^{\langle t+1\rangle}_{\langle t\rangle,1}, w_1^{\langle t+1\rangle}; \theta^{\star}_{\langle t\rangle,1}) / m(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta^{\star}_{\langle t\rangle,1}) \\
U_2 &= m(\|\boldsymbol{\theta}^{\langle t+1\rangle}\|, w_1^{\langle t+1\rangle}; \|\boldsymbol{\theta}^{\star}\|) / m(\theta^{\langle t+1\rangle}_{\langle t\rangle,1}, w_1^{\langle t\rangle}; \theta^{\star}_{\langle t\rangle,1}) \\
U_3 &= m(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta^{\star}_{\langle t\rangle,1}) / m(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \|\boldsymbol{\theta}^{\star}\|).
\end{aligned}$$

For any $\delta_0 > 0$, we have after finite iterations t_1, $\theta^{\star}_{\langle t_1\rangle,1}$ will stay in the δ_0-neighborhood around $\|\boldsymbol{\theta}^{\star}\|$. Hence, consider $t > t_1$, note that on the following com-

pact set $\mathcal{S}'$:

$$\mathcal{S}' := \left\{ w_1^{\langle t \rangle} \in [c_3, c_4], \|\boldsymbol{\theta}^{\langle t \rangle}\| \in [c_1, c_2], \theta^{\star}_{\langle t \rangle, 1} \in [\|\boldsymbol{\theta}^{\star}\| - \delta_0, \|\boldsymbol{\theta}^{\star}\|] \right\} - \left\{ (\|\boldsymbol{\theta}^{\langle t \rangle}\| - \|\boldsymbol{\theta}^{\star}\|)^2 + (w_1^{\langle t \rangle} - w_1^{\star})^2 < 4\delta_0^2 \right\}. \tag{2.87}$$

we have $U_1 < 1$, therefore, we can find constant $\rho_1 < 1$ such that $U_1 \leq \rho_1$ on $\mathcal{S}'$. Further, we know there exists a constant c' such that $\max(U_2, U_3) \leq (1 + c' \cdot |\beta^{\langle t \rangle}|)$ on this compact set $\mathcal{S}'$ since $\theta^{\star}_{\langle t \rangle, 1} = \cos \beta^{\langle t \rangle} \cdot \|\boldsymbol{\theta}^{\star}\|$ and $\theta^{\langle t+1 \rangle}_{\langle t \rangle, 1} = \cos \alpha^{\langle t \rangle} \cdot \|\boldsymbol{\theta}^{\langle t+1 \rangle}\|$ with $|\alpha^{\langle t \rangle}| < |\beta^{\langle t \rangle}|$. Hence for large enough t_2, there exists $\rho_2 < 1$ such that for any $t > t_2$ and point $(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle})$ in $\mathcal{S}'$, we have

$$\frac{m(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle}; \|\boldsymbol{\theta}^{\star}\|)}{m(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle}; \|\boldsymbol{\theta}^{\star}\|)} = U_1 \cdot U_2 \cdot U_3 \leq \rho_2 < 1.$$

Hence, we have either $m(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle}; \|\boldsymbol{\theta}^{\star}\|)$ is strictly decreasing at rate ρ_2 or $(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle})$ was in the $2\delta_0$-neighborhood around $(\|\boldsymbol{\theta}^{\star}\|, w_1^{\star})$ and therefore by the analysis in Lemma 2.14, there exists constant $c'' > 0$ and $c''' > 0$ such that

$$m(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle}; \|\boldsymbol{\theta}^{\star}\|) < (1 + c'' \cdot |\beta^{\langle t \rangle}|) \cdot c''' \delta_0^2.$$

Either way, by arbitrary choice of δ_0, we know $m(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle}; \|\boldsymbol{\theta}^{\star}\|)$ converges to 0 which implies $\boldsymbol{\theta}^{\langle t \rangle}$ converges to $\boldsymbol{\theta}^{\star}$. Hence, finally, we just need to bound $w_1^{\langle t \rangle}$. Note that in the proof of Lemma 2.14, we used the following strategy to show that $w_1^{\langle t \rangle}$ is bounded away from 1:

- If $(\theta^{\langle 0 \rangle}, w_1^{\langle 0 \rangle}) \in R_5 \bigcup R_6$, within finite iterations t_0, $(\theta^{\langle t_0 \rangle}, w_1^{\langle t_0 \rangle})$ will reach the region $R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4$.

- When $(\theta^{\langle t_0\rangle}, w_1^{\langle t_0\rangle}) \in R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4$, by Equation (2.69) and Equation (2.70), we have for all $t \geq t_0$,

$$(\theta^{\langle t+1\rangle}, w_1^{\langle t+1\rangle}) \in D(\theta^{\langle t+1\rangle}, w_1^{\langle t+1\rangle}) \overset{(a)}{\subseteq} D(\theta^{\langle t\rangle}, w_1^{\langle t\rangle}) \subseteq \cdots \subseteq D(\theta^{\langle t_0\rangle}, w_1^{\langle t_0\rangle}). \quad (2.88)$$

  Hence, $w^{\langle t\rangle} \leq \max(w_1^{\langle t_0\rangle}, r^{-1}(\theta^{\langle t_0\rangle}))$.

However, in multi-dimsnional case, since we changed the true values $\theta_\star$ from $\theta^\star_{\langle t\rangle,1}$ to $\theta^\star_{\langle t+1\rangle,1}$ after each iteration, definition of R_5 and R_6 changes and relation (a) in Equation (2.88) does not hold anymore, namely,

$$D(\theta^{\langle t+1\rangle}_{\langle t\rangle,1}, w_1^{\langle t+1\rangle}; \theta^\star_{\langle t+1\rangle,1}) \not\subset D(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta^\star_{\langle t\rangle,1}).$$

Yet, we can have a quick remedy for this strategy. Note that since $\theta^\star_{\langle t\rangle,1} \to \|\boldsymbol{\theta}^\star\|$, our adjusted reference curve $r_{adj}(w_1; \theta^\star_{\langle t\rangle,1})$ also converges to $r_{adj}(w_1; \|\boldsymbol{\theta}^\star\|)$ uniformly for $w_1 \in [w_1^\star, 1]$. Hence, we can find $\delta' > 0$, $t' > 0$ such that we can perturb every $r_{adj}(w_1; \theta^\star_{\langle t\rangle,1})$ for $t > t'$ such that we have $\tilde{r}_{adj}(w_1; \theta^\star_{\langle t\rangle,1})$ satisfies C.1 and C.2 for true value $\theta_\star = \theta^\star_{\langle t\rangle,1}$ for all $t > t'$ with

$$\tilde{r}_{adj}(w_1; \theta_\star) \;=\; r_{adj}(w_1; \theta^\star_{\langle t'\rangle,1}), \quad \forall w_1 \in [1-\delta', 1], \theta_\star \in [\theta^\star_{\langle t'\rangle,1}, \|\boldsymbol{\theta}^\star\|],$$

and

$$\tilde{r}_{adj}(w_1; \theta_\star) \;=\; r(w_1; \theta_\star), \quad \forall w_1 \leq w_1^\star, \theta_\star \in [\theta^\star_{\langle t'\rangle,1}, \|\boldsymbol{\theta}^\star\|].$$

Hence, the region R_5 and R_6 are invariant for $\theta_\star \in [\theta^\star_{\langle t'\rangle,1}, \|\boldsymbol{\theta}^\star\|]$, and therefore with the same arguments made in the proof of Lemma 2.14, within finite iterations t'', we have

$$\|\boldsymbol{\theta}^{\langle t''\rangle}\| > \theta^\star_{\langle t'\rangle,1}(w_1^\star - w_2^\star),$$

in other words, $(\|\boldsymbol{\theta}^{\langle t'' \rangle}\|, w_1^{\langle t'' \rangle})$ lies in $R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4$ for any true value $\theta_\star \in [\theta^\star_{\langle t' \rangle,1}, \|\boldsymbol{\theta}^\star\|]$. Once the point $(\|\boldsymbol{\theta}^{\langle t'' \rangle}\|, w_1^{\langle t'' \rangle})$ lies in the region $R_1 \bigcup R_2 \bigcup R_3 \bigcup R_4$, we can bound every $(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle})$ for all $t \geq t''$ by the following union of two rectangles

$$D\left(\min\left(\tilde{r}_{adj}(1-\delta'), \|\boldsymbol{\theta}^{\langle t \rangle}\|\right), \min\left(c_3, r^{-1}(c_2; \theta^\star_{\langle t \rangle,1})\right); \|\boldsymbol{\theta}^\star\|\right) \bigcup D\left(c_2, \max(w_1^{\langle t \rangle}, 1-\delta'); \theta^\star_{\langle t \rangle,1}\right), \tag{2.89}$$

due to the fact that $(\theta^{\langle t+1 \rangle}_{\langle t \rangle,1}, w_1^{\langle t+1 \rangle}) \in D(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle}; \theta^\star_{\langle t \rangle,1})$ and $\|\boldsymbol{\theta}^{\langle t+1 \rangle}\| \leq c_2$. Denote the set defined in Equation (2.89) as $\mathcal{Q}(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle})$. Then, we can check that for any $(\theta, w_1) \in \mathcal{Q}(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle})$, we have $\mathcal{Q}(\theta, w_1) \subseteq \mathcal{Q}(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w_1^{\langle t \rangle})$. Therefore, we have $\mathcal{Q}(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle}) \subseteq \mathcal{Q}(\|\boldsymbol{\theta}^{\langle t \rangle}\|, w^{\langle t \rangle})$. Hence, by a chain of arguments starting from t'', we have

$$(\|\boldsymbol{\theta}^{\langle t+1 \rangle}\|, w_1^{\langle t+1 \rangle}) \in \mathcal{Q}(\|\boldsymbol{\theta}^{\langle t'' \rangle}\|, w_1^{\langle t'' \rangle}).$$

Hence, we have

$$w_1^{\langle t \rangle} \leq \max\left(\tilde{r}^{-1}_{adj}(\|\boldsymbol{\theta}^{\langle t'' \rangle}\|; \|\boldsymbol{\theta}^\star\|), 1-\delta', w_1^{\langle t'' \rangle}\right) \; < \; 1, \quad \forall t \geq t''.$$

2.4.4.3 Geometric convergence

Since we have shown that $(\boldsymbol{\theta}^{\langle t \rangle}, w^{\langle t \rangle})$ converges to $(\boldsymbol{\theta}^\star, w_1^\star)$, we just need to show an attraction basin around $(\boldsymbol{\theta}^\star, w_1^\star)$, and therefore, combining both, we know after a finite iteration T, we have geometric convergence. To show an attraction basin, let us consider the following two terms $\|\boldsymbol{\theta}^{\langle t+1 \rangle} - \boldsymbol{\theta}^\star\|$ and $|w_1^{\langle t+1 \rangle} - w_1^\star|$. Apply the sequence

of the coordinate system $\mathcal{A}$, then by Equation (2.85) and Equation (A.20), we have

$$\begin{aligned}
\|\boldsymbol{\theta}^{\langle t+1\rangle} - \boldsymbol{\theta}^\star\|^2 &= |\theta_{\langle t\rangle,1}^{\langle t+1\rangle} - \theta_{\langle t\rangle,1}^\star|^2 + |\theta_{\langle t\rangle,2}^{\langle t+1\rangle} - \theta_{\langle t\rangle,2}^\star|^2 \\
&= |G_\theta(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta_{\langle t\rangle,1}^\star, w_1^\star) - \theta_{\langle t\rangle,1}^\star|^2 + |\theta_{\langle t\rangle,2}^\star|^2 (1 - s(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta_{\langle t\rangle,1}^\star, w_1^\star))^2, \\
|w_1^{\langle t+1\rangle} - w_1^\star| &= |G_w(\|\boldsymbol{\theta}^{\langle t\rangle}\|, w_1^{\langle t\rangle}; \theta_{\langle t\rangle,1}^\star, w_1^\star) - w_1^\star| \qquad (2.90)
\end{aligned}$$

Hence, we just need to show that for all $\theta_\|^\star > 0$ and $w_1^\star \in (0,1)$, the eigenvalues of the Jacobian matrix of the following mapping:

$$(\theta, w_1) \mapsto (G_\theta(\theta, w_1; \theta_\|^\star, w_1^\star), G_w(\theta, w_1; \theta_\|^\star, w_1^\star)) \qquad (2.91)$$

are in $[0,1)$ at $(\theta, w_1) = (\theta_\|^\star, w_1^\star)$. Then, note that

$$G_\theta(\theta_\|^\star, w_1^\star; \theta_\|^\star, w_1^\star) = \theta_\|^\star \quad \text{and} \quad G_w(\theta_\|^\star, w_1^\star; \theta_\|^\star, w_1^\star) = w_1^\star.$$

Hence, by continuity of the Jacobian of the functions, there exists $\epsilon > 0$ and $\rho < 1$ such that as long as $\theta, \theta_\|^\star \in [\|\boldsymbol{\theta}^\star\| - \epsilon, \|\boldsymbol{\theta}^\star\| + \epsilon]$ and $w_1 \in [w_1^\star - \epsilon, w_1^\star + \epsilon]$, we have

$$(G_\theta(\theta, w_1; \theta_\|^\star, w_1^\star) - \theta_\|^\star)^2 + (G_w(\theta, w_1; \theta_\|^\star, w_1^\star) - w_1^\star)^2 \le \rho\left((\theta - \theta_\|^\star)^2 + (w_1 - w_1^\star)^2\right).$$

Further, by Equation (A.23), we know function $s(\theta, w_1; \theta_\|^\star, w_1^\star)$ is positive on $\theta, \theta_\|^\star \in [\|\boldsymbol{\theta}^\star\| - \epsilon, \|\boldsymbol{\theta}^\star\| + \epsilon]$ and $w_1 \in [w_1^\star - \epsilon, w_1^\star + \epsilon]$. Hence, there exists constant ρ' such that

$$(1 - s(\theta, w_1; \theta_\|^\star, w_1^\star))^2 \le \rho', \quad \forall \theta, \theta_\|^\star \in [\|\boldsymbol{\theta}^\star\| - \epsilon, \|\boldsymbol{\theta}^\star\| + \epsilon], w_1 \in [w_1^\star - \epsilon, w_1^\star + \epsilon].$$

Hence, plug in Equation (2.90), we have if $\|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1} \in [\|\boldsymbol{\theta}^{\star}\| - \epsilon, \|\boldsymbol{\theta}^{\star}\| + \epsilon]$ and $w_1^{\langle t\rangle} \in [w_1^{\star} - \epsilon, w_1^{\star} + \epsilon]$, then

$$\begin{aligned}\|\boldsymbol{\theta}^{\langle t+1\rangle} - \boldsymbol{\theta}^{\star}\|^2 + |w_1^{\langle t+1\rangle} - w_1^{\star}|^2 &\leq \rho\left((\|\boldsymbol{\theta}^{\langle t\rangle}\| - \theta^{\star}_{\langle t\rangle,1})^2 + (w_1^{\langle t\rangle} - w_1^{\star})^2\right) + \rho'|\theta^{\star}_{\langle t\rangle,2}|^2 \\ &\leq \max(\rho,\rho')\left(\|\boldsymbol{\theta}^{\langle t\rangle} - \boldsymbol{\theta}^{\star}\|^2 + (w_1^{\langle t\rangle} - w_1^{\star})^2\right).\end{aligned}$$

Hence, by triangle inequality, we know once $\|\boldsymbol{\theta}^{\langle t\rangle} - \boldsymbol{\theta}^{\star}\| \leq \epsilon$ and $|w_1^{\langle t\rangle} - w_1^{\star}| \leq \epsilon$, we have $(\boldsymbol{\theta}^{\langle t\rangle}, w_1^{\langle t\rangle})$ geometrically converges towards $(\boldsymbol{\theta}^{\star}, w_1^{\star})$. Further, the first iteration to reach the attraction basin is guaranteed by the geometric convergence of the angle $\beta^{\langle t\rangle}$ and geometric convergence of the area function $m(\theta, w)$ on $\mathcal{S}'$ defined in Equation (2.87) for $\delta_0 = \epsilon/4$.

Next, we will show that for all $\theta^{\star}_{\|} > 0$ and $w_1^{\star} \in (0,1)$, the eigenvalues of the Jacobian matrix of the mapping defined in Equation (2.91) at $(\theta, w_1) = (\theta^{\star}_{\|}, w_1^{\star})$ are in $[0,1)$. Note that this Jacobian matrix at $(\theta, w_1) = (\theta^{\star}_{\|}, w_1^{\star})$ is the following:

$$J = \begin{bmatrix} \underbrace{\int \frac{4w_1^{\star}w_2^{\star}y^2}{w_1^{\star}e^{y\theta^{\star}_{\|}} + w_2^{\star}e^{-y\theta^{\star}_{\|}}}\phi(y)e^{-\frac{(\theta^{\star}_{\|})^2}{2}}\mathrm{d}y}_{J_{11}} & \underbrace{\int \frac{2y}{w_1^{\star}e^{y\theta^{\star}_{\|}} + w_2^{\star}e^{-y\theta^{\star}_{\|}}}\phi(y)e^{-\frac{(\theta^{\star}_{\|})^2}{2}}\mathrm{d}y}_{J_{12}} \\ \underbrace{\int \frac{2w_1^{\star}w_2^{\star}y}{w_1^{\star}e^{y\theta^{\star}_{\|}} + w_2^{\star}e^{-y\theta^{\star}_{\|}}}\phi(y)e^{-\frac{(\theta^{\star}_{\|})^2}{2}}\mathrm{d}y}_{J_{21}} & \underbrace{\int \frac{1}{w_1^{\star}e^{y\theta^{\star}_{\|}} + w_2^{\star}e^{-y\theta^{\star}_{\|}}}\phi(y)e^{-\frac{(\theta^{\star}_{\|})^2}{2}}\mathrm{d}y}_{J_{22}} \end{bmatrix}.$$

Then the two eigenvalues of J should be the two solutions of the following equation:

$$q(\lambda) = \lambda^2 - \lambda(J_{11} + J_{22}) + J_{11}J_{22} - J_{12}J_{21} = 0.$$

Note that, by Cauchy-Schwarz inequality, we know $\det(J) = J_{11}J_{22} - J_{12}J_{21} \geq 0$ and

therefore $q(0) \geq 0$. Also note that

$$q(J_{22}) \;=\; -J_{22}^2 - J_{12}J_{21} \;\leq\; 0,$$

and

$$\begin{aligned}
0 \;<\; J_{22} \;&=\; \int_{y\geq 0} \frac{e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star}}{w_1^\star w_2^\star (e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2 + 1} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \\
&=\; \int_{y\geq 0} (e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star}) \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \\
&\qquad\qquad - \int_{y\geq 0} \frac{w_1^\star w_2^\star (e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star})(e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2}{w_1^\star w_2^\star (e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2 + 1} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \\
&=\; 1 - \int_{y\geq 0} \frac{w_1^\star w_2^\star (e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star})(e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2}{w_1^\star w_2^\star (e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2 + 1} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \\
&\leq\; 1. \qquad (2.92)
\end{aligned}$$

Hence, we just need to show $q(1) > 0$, then the two solutions of $q(\lambda) = 0$ should stay in $[0, 1)$. Note that

$$\begin{aligned}
J_{11} \;&=\; \int_{y\geq 0} \frac{4 w_1^\star w_2^\star (e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star}) y^2}{w_1^\star w_2^\star (e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2 + 1} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \\
&=\; \int_{y\geq 0} \frac{4y^2}{e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star}} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \\
&\qquad\qquad - \int_{y\geq 0} \frac{4(w_1^\star - w_2^\star)^2 y^2}{(e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star})(w_1^\star w_2^\star (e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2 + 1)} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \\
&<\; 1 - \int_{y\geq 0} \frac{4(w_1^\star - w_2^\star)^2 y^2}{(e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star})(w_1^\star w_2^\star (e^{y\theta_{\|}^\star} - e^{-y\theta_{\|}^\star})^2 + 1)} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y, \qquad (2.93)
\end{aligned}$$

where the last inequality holds due to the fact that

$$\int_{y\geq 0} \frac{4y^2}{e^{y\theta_{\|}^\star} + e^{-y\theta_{\|}^\star}} \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \;\leq\; \int_{y\geq 0} 2y^2 \phi(y) e^{-\frac{(\theta_{\|}^\star)^2}{2}} \mathrm{d}y \;=\; e^{-\frac{(\theta_{\|}^\star)^2}{2}}.$$

Combine Equation (2.92) and Equation (2.93), we have

$$\begin{aligned} q(1) &= (1-J_{11})(1-J_{22}) - J_{12}J_{21} \\ &> \int_{y\geq 0} \frac{4(w_1^\star - w_2^\star)^2 y^2}{(e^{y\theta_\parallel^\star} + e^{-y\theta_\parallel^\star})(w_1^\star w_2^\star (e^{y\theta_\parallel^\star} - e^{-y\theta_\parallel^\star})^2 + 1)} \phi(y) e^{-\frac{(\theta_\parallel^\star)^2}{2}} \mathrm{d}y \\ &\quad \times \int_{y\geq 0} \frac{w_1^\star w_2^\star (e^{y\theta_\parallel^\star} + e^{-y\theta_\parallel^\star})(e^{y\theta_\parallel^\star} - e^{-y\theta_\parallel^\star})^2}{w_1^\star w_2^\star (e^{y\theta_\parallel^\star} - e^{-y\theta_\parallel^\star})^2 + 1} \phi(y) e^{-\frac{(\theta_\parallel^\star)^2}{2}} \mathrm{d}y \\ &\quad -4 w_1^\star w_2^\star (w_1^\star - w_2^\star)^2 \int_{y\geq 0} \left(\frac{(e^{y\theta_\parallel^\star} - e^{-y\theta_\parallel^\star}) y}{w_1^\star w_2^\star (e^{y\theta_\parallel^\star} - e^{-y\theta_\parallel^\star})^2 + 1} \phi(y) e^{-\frac{(\theta_\parallel^\star)^2}{2}} \mathrm{d}y \right)^2 \\ &\geq 0, \end{aligned}$$

where the last inequality holds due to Cauchy-Schwarz inequality. Hence, we have $q(1) > 0$ and this completes our proof for geometric convergence of the EM estimates.

2.5 Proof for Sample-based EM's results

It is straightforward to show that Sample-based EM also has the property of Lemma 2.3. Therefore, without loss of generality, we assume $\boldsymbol{\Sigma} = \boldsymbol{I}$.

2.5.1 Proof of Theorem 2.5

Let $\hat{\boldsymbol{a}}^{\langle t \rangle} = \frac{\hat{\boldsymbol{\mu}}_1^{\langle t \rangle} + \hat{\boldsymbol{\mu}}_2^{\langle t \rangle}}{2}$ and $\hat{\boldsymbol{\theta}}^{\langle t \rangle} = \frac{\hat{\boldsymbol{\mu}}_2^{\langle t \rangle} - \hat{\boldsymbol{\mu}}_1^{\langle t \rangle}}{2}$. Then the iteration functions based on $(\hat{\boldsymbol{a}}^{\langle t \rangle}, \hat{\boldsymbol{\theta}}^{\langle t \rangle})$ are the following:

$$\hat{\boldsymbol{a}}^{\langle t+1 \rangle} = \frac{\hat{\boldsymbol{q}}^{\langle t+1 \rangle}(1 - 2\hat{\mathsf{p}}^{\langle t+1 \rangle})}{2\hat{\mathsf{p}}^{\langle t+1 \rangle}(1 - \hat{\mathsf{p}}^{\langle t+1 \rangle})} + \frac{\bar{\boldsymbol{y}}}{2(1 - \hat{\mathsf{p}}^{\langle t+1 \rangle})}, \tag{2.94}$$

$$\hat{\boldsymbol{\theta}}^{\langle t+1 \rangle} = \frac{\hat{\boldsymbol{q}}^{\langle t+1 \rangle}}{2\hat{\mathsf{p}}^{\langle t+1 \rangle}(1 - \hat{\mathsf{p}}^{\langle t+1 \rangle})} - \frac{\bar{\boldsymbol{y}}}{2(1 - \hat{\mathsf{p}}^{\langle t+1 \rangle})}. \tag{2.95}$$

where

$$\begin{aligned}
\bar{\boldsymbol{y}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i, \\
\hat{\boldsymbol{q}}^{\langle t+1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{w}_d(\boldsymbol{y}_i-\hat{\boldsymbol{a}}^{\langle t\rangle},\hat{\boldsymbol{\theta}}^{\langle t\rangle})\boldsymbol{y}_i, \qquad (2.96)\\
\hat{\mathsf{p}}^{\langle t+1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{w}_d(\boldsymbol{y}_i-\hat{\boldsymbol{a}}^{\langle t\rangle},\hat{\boldsymbol{\theta}}^{\langle t\rangle}). \qquad (2.97)
\end{aligned}$$

Therefore $\hat{\boldsymbol{q}}^{\langle t\rangle}$ and $\hat{\mathsf{p}}^{\langle t\rangle}$ are the empirical versions of $\boldsymbol{\gamma}^{\langle t\rangle}$ and $\mathsf{p}^{\langle t\rangle}$ respectively. Our first goal is, for each $i\in\{1,2\}$, to compare the Population EM sequence $(\boldsymbol{\mu}_i^{\langle t\rangle})_{t\geq 0}$ to the Sample-based EM sequence $(\hat{\boldsymbol{\mu}}_i^{\langle t\rangle})_{t\geq 0}$, provided that the initial values $\boldsymbol{\mu}_i^{\langle 0\rangle}$ and $\hat{\boldsymbol{\mu}}_i^{\langle 0\rangle}$ are the same. We prove that

$$\hat{\boldsymbol{a}}^{\langle t\rangle}\to\boldsymbol{a}^{\langle t\rangle} \text{ in probability} \quad \text{and} \quad \hat{\boldsymbol{\theta}}^{\langle t\rangle}\to\boldsymbol{\theta}^{\langle t\rangle} \text{ in probability}, \quad \text{as } n\to\infty. \qquad (2.98)$$

We prove by induction. For $t=0$, it is clear that Equation (2.98) holds because both Population EM and Sample-based EM start with the same initialization. For $t=1$, by Weak Large Law Numbers (WLLN), we have

$$\begin{aligned}
\hat{\boldsymbol{q}}^{\langle 1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{w}_d(\boldsymbol{y}_i-\hat{\boldsymbol{a}}^{\langle 0\rangle},\hat{\boldsymbol{\theta}}^{\langle 0\rangle})\boldsymbol{y}_i \xrightarrow{p} \mathbb{E}\mathsf{w}_d(\boldsymbol{y}-\hat{\boldsymbol{a}}^{\langle 0\rangle},\hat{\boldsymbol{\theta}}^{\langle 0\rangle})y \;=\; \boldsymbol{\gamma}^{\langle 1\rangle},\\
\hat{\mathsf{p}}^{\langle 1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{w}_d(\boldsymbol{y}_i-\hat{\boldsymbol{a}}^{\langle 0\rangle},\hat{\boldsymbol{\theta}}^{\langle 0\rangle}) \xrightarrow{p} \mathbb{E}\mathsf{w}_d(\boldsymbol{y}-\hat{\boldsymbol{a}}^{\langle 0\rangle},\hat{\boldsymbol{\theta}}^{\langle 0\rangle}) \;=\; \mathsf{p}^{\langle 1\rangle},\\
\bar{\boldsymbol{y}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i \xrightarrow{p} \mathbb{E}\boldsymbol{y} \;=\; \mathbf{0}.
\end{aligned}$$

Since $\mathsf{p}^{\langle 1\rangle} \in (0,1)$, by employing the continuous mapping theorem, we have

$$\begin{aligned}
\hat{\boldsymbol{a}}^{\langle 1\rangle} &= \frac{\hat{\boldsymbol{q}}^{\langle 1\rangle}(1-2\hat{\mathsf{p}}^{\langle 1\rangle})}{2\hat{\mathsf{p}}^{\langle 1\rangle}(1-\hat{\mathsf{p}}^{\langle 1\rangle})} + \frac{\bar{\boldsymbol{y}}}{2(1-\hat{\mathsf{p}}^{\langle 1\rangle})} \;\rightarrow\; \frac{\boldsymbol{\gamma}^{\langle 1\rangle}(1-2\mathsf{p}^{\langle 1\rangle})}{2\mathsf{p}^{\langle 1\rangle}(1-\mathsf{p}^{\langle 1\rangle})} = \boldsymbol{a}^{\langle 1\rangle} \text{ in probability,} \\
\hat{\boldsymbol{\theta}}^{\langle 1\rangle} &= \frac{\hat{\boldsymbol{q}}^{\langle 1\rangle}}{2\hat{\mathsf{p}}^{\langle 1\rangle}(1-\hat{\mathsf{p}}^{\langle 1\rangle})} + \frac{\bar{\boldsymbol{y}}}{2(1-\hat{\mathsf{p}}^{\langle 1\rangle})} \;\rightarrow\; \frac{\boldsymbol{\gamma}^{\langle 1\rangle}}{2\mathsf{p}^{\langle 1\rangle}(1-\mathsf{p}^{\langle 1\rangle})} \;=\; \boldsymbol{\theta}^{\langle 1\rangle} \text{ in probability.}
\end{aligned}$$

Therefore Equation (2.98) holds for $t = 1$. Now we assume that Equation (2.98) holds for $t \geq 1$, and our goal is to prove it for $t+1$. Note that

$$\begin{aligned}
\left\| \frac{\partial \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta)}{\partial \boldsymbol{x}_a} \right\| &= \left\| -\frac{2\boldsymbol{x}_\theta}{(e^{\langle \boldsymbol{y}, \boldsymbol{x}_\theta\rangle - \langle \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle} + e^{-\langle \boldsymbol{y}, \boldsymbol{x}_\theta\rangle + \langle \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle})^2} \right\| \\
&\leq \frac{\|\boldsymbol{x}_\theta\|}{2}, \\
\left\| \frac{\partial \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta)}{\partial \boldsymbol{x}_\theta} \right\| &= \left\| \frac{2(\boldsymbol{y} - \boldsymbol{x}_a)}{(e^{\langle \boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle} + e^{-\langle \boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle})^2} \right\| \\
&\leq \left\| \frac{\boldsymbol{y} - \boldsymbol{x}_a}{2} \right\| \\
&\leq \frac{\|\boldsymbol{y}\| + \|\boldsymbol{x}_a\|}{2}.
\end{aligned}$$

Therefore we have

$$
\begin{aligned}
&|\mathsf{p}^{\langle t+1\rangle} - \hat{\mathsf{p}}^{\langle t+1\rangle}| = |\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})| \\
\leq\ & \left|\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\right| \\
&+ \left|\frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})\right| \\
\leq\ & \left|\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\right| \\
&+ \left|\frac{1}{n}\sum_{i=1}^{n} \left(\frac{\|\boldsymbol{\theta}_\xi^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{a}}^{\langle t\rangle} - \boldsymbol{a}^{\langle t\rangle}\| + \frac{\|\boldsymbol{y}_i\| + \|\boldsymbol{a}_\xi^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\|\right)\right| \\
\leq\ & \left|\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\right| + \frac{1}{2}\frac{\sum_{i=1}^{n}\|\boldsymbol{y}_i\|}{n}\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| \\
&+ \left(\frac{\|\boldsymbol{\theta}_\xi^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{a}}^{\langle t\rangle} - \boldsymbol{a}^{\langle t\rangle}\| + \frac{\|\boldsymbol{a}_\xi^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\|\right),
\end{aligned}
$$

where

$$
\boldsymbol{a}_\xi^{\langle t\rangle} = \xi\boldsymbol{a}^{\langle t\rangle} + (1-\xi)\hat{\boldsymbol{a}}^{\langle t\rangle}, \quad \text{and} \quad \boldsymbol{\theta}_\xi^{\langle t\rangle} = \xi\boldsymbol{\theta}^{\langle t\rangle} + (1-\xi)\hat{\boldsymbol{\theta}}^{\langle t\rangle}, \text{ for some } \xi \in [0,1].
$$

By WLLN, induction assumption and

$$
\|\boldsymbol{a}_\xi^{\langle t\rangle}\| \leq 2\|\boldsymbol{a}^{\langle t\rangle}\| + \|\boldsymbol{a}^{\langle t\rangle} - \hat{\boldsymbol{a}}^{\langle t\rangle}\|, \quad \text{and} \quad \|\boldsymbol{\theta}_\xi^{\langle t\rangle}\| \leq 2\|\boldsymbol{\theta}^{\langle t\rangle}\| + \|\boldsymbol{\theta}^{\langle t\rangle} - \hat{\boldsymbol{\theta}}^{\langle t\rangle}\|,
$$

we have

$$\begin{aligned}
|\mathsf{p}^{\langle t+1\rangle} - \hat{\mathsf{p}}^{\langle t+1\rangle}| &\leq |\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})| \\
&\quad + \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\| + \|\boldsymbol{\theta}^{\langle t\rangle} - \hat{\boldsymbol{\theta}}^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{a}}^{\langle t\rangle} - \boldsymbol{a}^{\langle t\rangle}\| \\
&\quad + \frac{2\|\boldsymbol{a}^{\langle t\rangle}\| + \|\boldsymbol{a}^{\langle t\rangle} - \hat{\boldsymbol{a}}^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| + \frac{1}{2}\frac{\sum_{i=1}^{n}\|\boldsymbol{y}_i\|}{n}\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| \\
&\to 0 \text{ in probability.}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\|\boldsymbol{\gamma}^{\langle t+1\rangle} - \hat{\boldsymbol{q}}^{\langle t+1\rangle}\| &= \|\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})\boldsymbol{y}_i\| \\
&\leq \|\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y}_i\| \\
&\quad + \|\frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})y_i - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})\boldsymbol{y}_i\| \\
&\leq \|\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y}_i\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n}(\frac{\|\boldsymbol{\theta}_\xi^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{a}}^{\langle t\rangle} - \boldsymbol{a}^{\langle t\rangle}\| + \frac{\|\boldsymbol{y}_i\| + \|\boldsymbol{a}_\xi^{\langle t\rangle}\|}{2}\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\|)\boldsymbol{y}_i\right\| \\
&\leq \left\|\mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} - \frac{1}{n}\sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y}_i\right\| \\
&\quad + \left\|\frac{1}{2n}\sum_{i=1}^{n}\|\boldsymbol{y}_i\|\boldsymbol{y}_i\right\|\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| + \Big(\|\boldsymbol{a}^{\langle t\rangle}\|\|\boldsymbol{\theta}^{\langle t\rangle} - \hat{\boldsymbol{\theta}}^{\langle t\rangle}\| + \|\boldsymbol{\theta}^{\langle t\rangle}\|\|\boldsymbol{a}^{\langle t\rangle} - \hat{\boldsymbol{a}}^{\langle t\rangle}\| \\
&\quad + \|\boldsymbol{a}^{\langle t\rangle} - \hat{\boldsymbol{a}}^{\langle t\rangle}\|\|\boldsymbol{\theta}^{\langle t\rangle} - \hat{\boldsymbol{\theta}}^{\langle t\rangle}\|\Big)\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\|.
\end{aligned}$$

By WLLN and induction assumption, we have

$$\|\boldsymbol{\gamma}^{\langle t+1\rangle} - \hat{\boldsymbol{q}}^{\langle t+1\rangle}\| \to 0 \text{ in probability.}$$

Therefore with $\mathsf{p}^{\langle t+1\rangle} \in (0,1)$, we have

$$\begin{aligned}
\hat{\boldsymbol{a}}^{\langle t+1\rangle} &= \frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}(1-2\hat{\mathsf{p}}^{\langle t+1\rangle})}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} + \frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \\
&\to \frac{\boldsymbol{\gamma}^{\langle t+1\rangle}(1-2\mathsf{p}^{\langle t+1\rangle})}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})} = \boldsymbol{a}^{\langle t+1\rangle} \text{ in probability,} \\
\hat{\boldsymbol{\theta}}^{\langle t+1\rangle} &= \frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} + \frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \\
&\to \frac{\boldsymbol{\gamma}^{\langle t+1\rangle}}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})} = \boldsymbol{\theta}^{\langle t+1\rangle} \text{ in probability.}
\end{aligned}$$

Hence Equation (2.98) holds for $t+1$. With induction, we completes the proof of this lemma.

2.5.2 Proof of Theorem 2.6

The main idea of the proof is simple. We first show that if we initialize Sample-based EM in a way that $\hat{\boldsymbol{a}}^{\langle 0\rangle}$ is small enough and $\hat{\boldsymbol{\theta}}^{\langle 0\rangle}$ is in small neighborhood of $\boldsymbol{\theta}^\star$, then the sampled based EM will converge to a point whose distance from $\boldsymbol{\theta}^\star$ is $O(\sqrt{d/n})$ with probability converging to 1 as $n \to \infty$. Let's call this neighborhood of $(\boldsymbol{a}, \boldsymbol{b})$, $\mathcal{N}_{\mathbf{0},\boldsymbol{\theta}^\star}$.

According to Theorem 2.2, we know that Population EM converges to the true parameter under quite general initialization. Hence, there exists an iteration T_0 at which the estimate of Population EM is in $\mathcal{N}_{\mathbf{0},\boldsymbol{\theta}^\star}$. We know from Theorem 2.5 that at iteration T_0, $\hat{\boldsymbol{a}}^{\langle T_0\rangle} \to \boldsymbol{a}^{\langle T_0\rangle}$ and $\hat{\boldsymbol{\theta}}^{\langle T_0\rangle} \to \boldsymbol{\theta}^{\langle T_0\rangle}$ in probability. Hence, with probability converging to 1, $(\hat{\boldsymbol{a}}^{\langle T_0\rangle}, \hat{\boldsymbol{\theta}}^{\langle T_0\rangle}) \in \mathcal{N}_{\mathbf{0},\boldsymbol{\theta}^\star}$, and hence $(\hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})$ converge to a point that is at a distance $O(\sqrt{d/n})$ from $(\mathbf{0}, \boldsymbol{\theta}^\star)$. In other words, if $\hat{\boldsymbol{a}}^\infty$ and $\hat{\boldsymbol{\theta}}^\infty$ the limiting

estimates, then

$$\|\hat{\boldsymbol{a}}^{\infty}\| = O(\sqrt{d/n}),$$
$$\|\hat{\boldsymbol{\theta}}^{\infty} - \boldsymbol{\theta}^{\star}\| = O(\sqrt{d/n}),$$

with probability converging to 1, which is equivalent to what we wanted to prove.

As is clear from the above discussion, the only challenging part is to prove that if $(\hat{\boldsymbol{a}}^{\langle 0\rangle}, \hat{\boldsymbol{\theta}}^{\langle 0\rangle})$ is in small neighborhood of $(\mathbf{0}, \boldsymbol{\theta}^{\star})$, then the sampled-based EM will converge to a point whose distance from $(\mathbf{0}, \boldsymbol{\theta}^{\star})$ is $O(\sqrt{d/n})$. The proof of this fact is our main goal in the rest of this proof.

We remind the reader that according to Theorems 2.2 the estimates of Population EM satisfy the following equations (if initialized properly):

$$\begin{aligned} \|\boldsymbol{a}^{\langle t\rangle}\| &\rightarrow 0, \\ \|\boldsymbol{\theta}^{\langle t\rangle} - \boldsymbol{\theta}^{\star}\| &\rightarrow 0, \end{aligned} \tag{2.99}$$

Also, we know from the arguments provided in the proof of Theorem 2.5 that $\hat{\boldsymbol{a}}^{\langle t\rangle}$ and $\hat{\boldsymbol{\theta}}^{\langle t\rangle}$ converge to $\boldsymbol{a}^{\langle t\rangle}$ and $\boldsymbol{\theta}^{\langle t\rangle}$ in probability. Hence, we expect to have a similar equations for $\hat{\boldsymbol{a}}^{\langle t\rangle}$ and $\hat{\boldsymbol{\theta}}^{\langle t\rangle}$, except for probably an error term that will vanish as $n \rightarrow \infty$. The only issue that may happen is that the errors that are introduced in each iteration may accumulate and will let to a non-vanishing error for $t \rightarrow \infty$. Our first lemma shows that this does not happen.

Lemma 2.19. Suppose that there exist $\kappa_a \in (0,1), \kappa_\theta \in (0,1)$ and $c_\theta > 0$ such that

for all $t' \geq 1$, we have

$$\begin{aligned} \|\hat{\boldsymbol{a}}^{\langle t' \rangle}\| &\leq \kappa_a \|\hat{\boldsymbol{a}}^{\langle t'-1 \rangle}\| + \epsilon_a, & (2.100) \\ \|\hat{\boldsymbol{\theta}}^{\langle t' \rangle} - \boldsymbol{\theta}^\star\| &\leq \kappa_\theta \|\hat{\boldsymbol{\theta}}^{\langle t'-1 \rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_\theta \|\hat{\boldsymbol{a}}^{\langle t'-1 \rangle}\|} + \epsilon_\theta, & (2.101) \end{aligned}$$

for some $\epsilon_a, \epsilon_\theta > 0$. Then we have $\forall t \geq 0$,

$$\begin{aligned} \|\hat{\boldsymbol{a}}^{\langle t \rangle}\| &\leq (\kappa_a)^t \|\hat{\boldsymbol{a}}^{\langle 0 \rangle}\| + \frac{1}{1-\kappa_a}\epsilon_a, & (2.102) \\ \|\hat{\boldsymbol{\theta}}^{\langle t \rangle} - \boldsymbol{\theta}^\star\| &\leq (\kappa_\theta)^t \|\hat{\boldsymbol{\theta}}^{\langle 0 \rangle} - \boldsymbol{\theta}^\star\| + t\sqrt{c_\theta \|\hat{\boldsymbol{a}}^{\langle 0 \rangle}\|}(\max\{\sqrt{\kappa_a}, \kappa_\theta\})^t \\ &\quad + \frac{1}{1-\kappa_\theta}\sqrt{\frac{c_\theta}{1-\kappa_a}\epsilon_a} + \frac{1}{1-\kappa_\theta}\epsilon_\theta & (2.103) \end{aligned}$$

We prove this lemma in Appendix A.2.1.1.

According to Lemma 2.19 as long as the errors that are introduced in each iteration are bounded by ϵ_a and ϵ_θ, the overall error will also remain bounded and are, in the worst case, proportional to $\sqrt{\epsilon_a}$ and ϵ_θ. Hence, if $\epsilon_a \to 0$ and $\epsilon_\theta \to 0$ as $n \to \infty$, the overall errors will go to zero too. Hence, proving that Equation (2.100) and Equation (2.101) hold for $\epsilon_a \to 0$ and $\epsilon \to 0$ will complete the proof Theorem 2.6. Indeed, the following lemma provides such claim.

Lemma 2.20. There exists constants $\kappa_a \in (\frac{\sqrt{3}}{2}, 1), \kappa_\theta \in (0, 1); c_\theta > 0$ and

$$\delta_a \in \left(0, \min\left\{1, \frac{\sqrt{3}}{2}\|\boldsymbol{\theta}^\star\|, \frac{(1-\kappa_\theta)^2(1-(\kappa_a)^2)\|\boldsymbol{\theta}^\star\|^2}{4c_\theta}\right\}\right)$$

only depending on $\boldsymbol{\theta}^\star$, such that if the initialization $(\hat{\boldsymbol{a}}^{\langle 0 \rangle}, \hat{\boldsymbol{\theta}}^{\langle 0 \rangle})$ satisfies

$$\|\hat{\boldsymbol{a}}^{\langle 0 \rangle}\| \leq \delta_a, \quad \text{and} \quad \|\hat{\boldsymbol{\theta}}^{\langle 0 \rangle} - \boldsymbol{\theta}^\star\| \leq \sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|,$$

then $\forall t \geq 0$, we have

$$\begin{aligned} \|\hat{\boldsymbol{a}}^{\langle t+1 \rangle}\| &\leq \kappa_a \|\hat{\boldsymbol{a}}^{\langle t \rangle}\| + \epsilon_a, \\ \|\hat{\boldsymbol{\theta}}^{\langle t+1 \rangle} - \boldsymbol{\theta}^\star\| &\leq \kappa_\theta \|\hat{\boldsymbol{\theta}}^{\langle t \rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_\theta \|\hat{\boldsymbol{a}}^{\langle t \rangle}\|} + \epsilon_\theta, \end{aligned}$$

with probability at least $1 - 3\delta$. The value of the other constants are the following

$$\begin{aligned} c_\theta &= 4(\|\boldsymbol{\theta}^\star\| + 2)\sqrt{\frac{3d + \ln(1/\delta)}{n}}, \\ C_\theta &= 3\|\boldsymbol{\theta}^\star\| c_\theta, \\ \epsilon_a = \epsilon_\theta &= \frac{9C_\theta + c_\theta}{\rho(2-\rho)} + \frac{12C_\theta}{\rho(2-\rho)} + \frac{c_\theta}{2-\rho}, \end{aligned} \tag{2.104}$$

where $\rho = \sup_{\|\boldsymbol{x}_a\| \leq 1, \|\boldsymbol{x}_\theta\| \leq \frac{3}{2}\|\boldsymbol{\theta}^\star\|} \max\{g_p(\boldsymbol{x}_a, \boldsymbol{x}_\theta, \boldsymbol{\theta}^\star), 1 - g_p(\boldsymbol{x}_a, \boldsymbol{x}_\theta, \boldsymbol{\theta}^\star)\} \in (0, 1)$. In addition, assume n is large enough to satisfy the following conditions:

$$\begin{aligned} C_\theta &< \frac{\rho}{2}, \\ \epsilon_a = \epsilon_\theta &\leq \min\left\{(1-\kappa_a)\delta_a, \frac{1}{2}(1-\kappa_\theta)\sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|\right\}. \end{aligned} \tag{2.105}$$

Proof. Note that

$$\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta) \triangleq \frac{e^{\langle \boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta \rangle}}{e^{\langle \boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta \rangle} + e^{-\langle \boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta \rangle}}.$$

We showed the following equations in Section 2.5.1:

$$\begin{aligned}\hat{\boldsymbol{a}}^{\langle t+1\rangle} &= \frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}(1-2\hat{\mathsf{p}}^{\langle t+1\rangle})}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})}+\frac{\bar{\boldsymbol{y}}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})},\\ \hat{\boldsymbol{\theta}}^{\langle t+1\rangle} &= \frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})}-\frac{\bar{\boldsymbol{y}}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})}.\end{aligned}$$

where

$$\begin{aligned}\bar{\boldsymbol{y}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i,\\ \hat{\boldsymbol{q}}^{\langle t+1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{w}_d(\boldsymbol{y}_i-\hat{\boldsymbol{a}}^{\langle t\rangle},\hat{\boldsymbol{\theta}}^{\langle t\rangle})\boldsymbol{y}_i,\\ \hat{\mathsf{p}}^{\langle t+1\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\mathsf{w}_d(\boldsymbol{y}_i-\hat{\boldsymbol{a}}^{\langle t\rangle},\hat{\boldsymbol{\theta}}^{\langle t\rangle}),\end{aligned}$$

We will show in Appendix A.2.1.2 that with probability at least $1-3\delta$, we have

$$\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\|\leq 4(\|\boldsymbol{\theta}^\star\|+2)\sqrt{\frac{3d+\ln(1/\delta)}{n}}=c_\theta, \tag{2.106}$$

$$\sup_{\substack{\|\boldsymbol{x}_\theta\|\leq\frac{3}{2}\|\boldsymbol{\theta}^\star\|,\\ \|\boldsymbol{x}_a\|\leq 1}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\mathbb{E}_Y\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{x}_a,\boldsymbol{x}_\theta)\right|\leq C_\theta, \tag{2.107}$$

$$\sup_{\substack{\|\boldsymbol{x}_\theta\|\leq\frac{3}{2}\|\boldsymbol{\theta}^\star\|,\\ \|\boldsymbol{x}_a\|\leq 1}}\left\|\frac{1}{n}\sum_{i=1}^{n}(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\boldsymbol{y}_i-\mathbb{E}(\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\boldsymbol{y}\right\|\leq\frac{9}{2}C_\theta. \tag{2.108}$$

Note that by setting $\delta=\frac{1}{n}$, we see that $c_\theta\to 0$, $C_\theta\to 0$, and $\delta\to 0$ simultaneously. In the rest of the proof we assume that Equation (2.106), Equation (2.107) and Equation

(2.108) hold. Let

$$\bar{\boldsymbol{\gamma}}^{\langle t+1\rangle} = \mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})\boldsymbol{y}, \quad \bar{\mathsf{p}}^{\langle t+1\rangle} = \mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle}),$$

and

$$\bar{\boldsymbol{a}}^{\langle t+1\rangle} = \frac{\bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}(1 - 2\bar{\mathsf{p}}^{\langle t+1\rangle})}{2\bar{\mathsf{p}}^{\langle t+1\rangle}(1 - \bar{\mathsf{p}}^{\langle t+1\rangle})}, \quad \bar{\boldsymbol{\theta}}^{\langle t+1\rangle} = \frac{\bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}}{2\bar{\mathsf{p}}^{\langle t+1\rangle}(1 - \bar{\mathsf{p}}^{\langle t+1\rangle})}.$$

The following lemma that will be proved in Appendix A.2.1.3 is a key step in our analysis:

Lemma 2.21. There exists $\kappa_a \in (0,1)$ such that if $\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| \leq \min\{\sqrt{1-(\kappa_a)^2}, \frac{1}{2}\}\|\boldsymbol{\theta}^\star\|$, then

$$\|\bar{\boldsymbol{a}}^{\langle t+1\rangle}\| \;\; \leq \;\; \kappa_a \|\hat{\boldsymbol{a}}^{\langle t\rangle}\|, \tag{2.109}$$

Furthermore, there exist $\delta'_a \in (0,1)$, $\kappa_\theta \in (0,1)$ and $c_\theta > 0$ such that if $\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| \in [0, \delta'_a]$, then

$$\|\bar{\boldsymbol{\theta}}^{\langle t+1\rangle} - \boldsymbol{\theta}^\star\| \;\; \leq \;\; \kappa_\theta \|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_b \|\hat{\boldsymbol{a}}^{\langle t\rangle}\|}. \tag{2.110}$$

Constant $\kappa_a, \kappa_\theta, \delta'_a$ and c_θ only depend on $\boldsymbol{\theta}^\star$.

The above equations provide connections between $(\bar{\boldsymbol{a}}^{\langle t+1\rangle}, \bar{\boldsymbol{\theta}}^{\langle t+1\rangle})$ and $(\hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})$. Next, we establish connection between $(\bar{\boldsymbol{a}}^{\langle t+1\rangle}, \bar{\boldsymbol{\theta}}^{\langle t+1\rangle})$ and $(\hat{\boldsymbol{a}}^{\langle t+1\rangle}, \hat{\boldsymbol{\theta}}^{\langle t+1\rangle})$. In the rest of the proof we assume that $\kappa_a \in (\sqrt{3}/2, 1)$. If κ_a is less than $\sqrt{3}/2$ we set it to $\sqrt{3}/2$. This is just for making notations simpler and has no specific technical reason.

Note that from Equation (2.94), we have

$$\begin{aligned}
\|\hat{\boldsymbol{a}}^{\langle t+1\rangle}\| &= \left\| \frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}(1-2\hat{\mathsf{p}}^{\langle t+1\rangle})}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} + \frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| \\
&\leq \left\| \frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}(1-2\hat{\mathsf{p}}^{\langle t+1\rangle})}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| + \left\| \frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| \\
&\leq \left| \frac{(1-2\hat{\mathsf{p}}^{\langle t+1\rangle})}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right| \|\hat{\boldsymbol{q}}^{\langle t+1\rangle} - \bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}\| + \left\| \frac{\bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}(1-2\bar{\mathsf{p}}^{\langle t+1\rangle})}{2\bar{\mathsf{p}}^{\langle t+1\rangle}(1-\bar{\mathsf{p}}^{\langle t+1\rangle})} \right\| + \left\| \frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| \\
&\quad + \left\| \frac{\bar{\boldsymbol{\theta}}^{\langle t+1\rangle}((\bar{\mathsf{p}}^{\langle t+1\rangle})^2 + (1-\bar{\mathsf{p}}^{\langle t+1\rangle})^2 - (1-2\bar{\mathsf{p}}^{\langle t+1\rangle})|\hat{\mathsf{p}}^{\langle t+1\rangle} - \bar{\mathsf{p}}^{\langle t+1\rangle}|)}{\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| |\hat{\mathsf{p}}^{\langle t+1\rangle} - \bar{\mathsf{p}}^{\langle t+1\rangle}| \\
&\leq \left| \frac{1}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right| \|\hat{\boldsymbol{q}}^{\langle t+1\rangle} - \bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}\| + \left\| \frac{3\bar{\boldsymbol{\theta}}^{\langle t+1\rangle}}{\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| |\hat{\mathsf{p}}^{\langle t+1\rangle} - \bar{\mathsf{p}}^{\langle t+1\rangle}| \\
&\quad + \|\bar{\boldsymbol{a}}^{\langle t+1\rangle}\| + \left\| \frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\|. \qquad (2.111)
\end{aligned}$$

Furthermore, from Equation (2.95) we have

$$\begin{aligned}
\|\hat{\boldsymbol{\theta}}^{\langle t+1\rangle} - \bar{\boldsymbol{\theta}}^{\langle t+1\rangle}\| &= \|\frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} - \frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} - \frac{\bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}}{2\bar{\mathsf{p}}^{\langle t+1\rangle}(1-\bar{\mathsf{p}}^{\langle t+1\rangle})}\| \\
&\leq \|\frac{\hat{\boldsymbol{q}}^{\langle t+1\rangle}}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} - \frac{\bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}}{2\bar{\mathsf{p}}^{\langle t+1\rangle}(1-\bar{\mathsf{p}}^{\langle t+1\rangle})}\| + \|\frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})}\| \\
&\leq \left| \frac{1}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right| \|\hat{\boldsymbol{q}}^{\langle t+1\rangle} - \bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}\| + \|\frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})}\| \\
&\quad + \left\| \frac{\bar{\boldsymbol{\theta}}^{\langle t+1\rangle}(1-2\bar{\mathsf{p}}^{\langle t+1\rangle} + |\hat{\mathsf{p}}^{\langle t+1\rangle} - \bar{\mathsf{p}}^{\langle t+1\rangle}|)}{\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| |\hat{\mathsf{p}}^{\langle t+1\rangle} - \bar{\mathsf{p}}^{\langle t+1\rangle}| \\
&\leq \left| \frac{1}{2\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right| \|\hat{\boldsymbol{q}}^{\langle t+1\rangle} - \bar{\boldsymbol{\gamma}}^{\langle t+1\rangle}\| \\
&\quad + \left\| \frac{3\bar{\boldsymbol{\theta}}^{\langle t+1\rangle}}{\hat{\mathsf{p}}^{\langle t+1\rangle}(1-\hat{\mathsf{p}}^{\langle t+1\rangle})} \right\| |\hat{\mathsf{p}}^{\langle t+1\rangle} - \bar{\mathsf{p}}^{\langle t+1\rangle}| + \|\frac{\bar{y}}{2(1-\hat{\mathsf{p}}^{\langle t+1\rangle})}\|.
\end{aligned} \qquad (2.112)$$

Suppose for the moment that $\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| \in [0,1]$ and $\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^{\star}\| \leq \frac{1}{2}\|\boldsymbol{\theta}^{\star}\|$. It is straightforward to use Equation (2.107) and Equation (2.105) and the definition of ρ in the

statement of Lemma 2.20 to prove

$$\hat{\mathsf{p}}^{\langle t+1\rangle} \in (\frac{\rho}{2}, 1-\frac{\rho}{2}). \tag{2.113}$$

By combining Equation (2.106)-Equation (2.108), Equation (2.111)), Equation (2.112), and Equation (2.113) we obtain

$$\begin{aligned} \|\hat{\boldsymbol{a}}^{\langle t+1\rangle}\| &\leq \|\bar{\boldsymbol{a}}^{\langle t+1\rangle}\| + \frac{9C_\theta + c_\theta}{\rho(2-\rho)} + \frac{12C_\theta}{\rho(2-\rho)} + \frac{c_\theta}{2-\rho} = \|\bar{\boldsymbol{a}}^{\langle t+1\rangle}\| + \epsilon_a, \\ \|\hat{\boldsymbol{\theta}}^{\langle t+1\rangle} - \bar{\boldsymbol{\theta}}^{\langle t+1\rangle}\| &\leq \frac{9C_\theta + c_\theta}{\rho(2-\rho)} + \frac{12C_\theta}{\rho(2-\rho)} + \frac{c_\theta}{2-\rho} = \epsilon_\theta, \end{aligned} \tag{2.114}$$

and hence $\|\hat{\boldsymbol{\theta}}^{\langle t+1\rangle} - \boldsymbol{\theta}^\star\| \leq \|\bar{\boldsymbol{\theta}}^{\langle t+1\rangle} - \boldsymbol{\theta}^\star\| + \epsilon_\theta$.

Now suppose that the assumptions of Lemma 2.21 hold, i.e., $\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| \in [0, \delta_a']$ and $\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| \leq \sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|$. Then Equation (2.114) implies that

$$\begin{aligned} \|\hat{\boldsymbol{a}}^{\langle t+1\rangle}\| &\leq \|\bar{\boldsymbol{a}}^{\langle t+1\rangle}\| + \epsilon_a \leq \kappa_a\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| + \epsilon_a, \\ \|\hat{\boldsymbol{\theta}}^{\langle t+1\rangle} - \boldsymbol{\theta}^\star\| &\leq \|\bar{\boldsymbol{\theta}}^{\langle t+1\rangle} - \boldsymbol{\theta}^\star\| + \epsilon_\theta \leq \kappa_\theta\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_b\|\hat{\boldsymbol{a}}^{\langle t\rangle}\|} + \epsilon_\theta. \end{aligned} \tag{2.115}$$

Note that Equation (2.115) is the result we claimed in Lemma 2.20. However, to obtain Equation (2.110), which is one of the main steps in deriving Equation (2.115) we have assumed that

$$\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| \in [0, \delta_a'] \quad \text{and} \quad \|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| \leq \sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|.$$

In order to prove the above equation holds for every t, we will prove an even stronger statement:

$$\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| \in [0, \delta_a] \quad \text{and} \quad \|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| \leq \sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|, \tag{2.116}$$

where $\delta_a = \min\{\delta'_a, \frac{(1-\kappa_\theta)^2(1-(\kappa_a)^2)\|\boldsymbol{\theta}^\star\|^2}{4c_\theta}\}$. We use induction to prove that Equation (2.116) holds $\forall t \geq 0$. By the assumptions of this Lemma, the initial estimates $(\hat{\boldsymbol{a}}^{\langle 0\rangle}, \hat{\boldsymbol{\theta}}^{\langle 0\rangle})$ satisfy Equation (2.116). Hence the base of the induction is true. Suppose Equation (2.116) holds for $t \geq 0$, then for $t+1$ Equation (2.115) holds. Hence all we need to prove is that

$$\kappa_a\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| + \epsilon_a \leq \delta_a,$$

and

$$\kappa_\theta\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_b\|\hat{\boldsymbol{a}}^{\langle t\rangle}\|} + \epsilon_\theta \leq \sqrt{1-(\kappa_a)^2}\}\|\boldsymbol{\theta}^\star\|. \tag{2.117}$$

For the first inequality, since the condition on n in Equation (2.105) ensure that $\epsilon_a \leq (1-\kappa_a)\delta_a$, together with induction assumption that $\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| \leq \delta_a$, we have

$$\|\hat{\boldsymbol{a}}^{\langle t+1\rangle}\| \leq \kappa_a\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| + \epsilon_a \leq \kappa_a\delta_a + (1-\kappa_a)\delta_a \leq \delta_a.$$

To prove Equation (2.117) note that the condition on n ensure that

$$\epsilon_\theta \leq \frac{1}{2}(1-\kappa_\theta)\sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|.$$

Also the condition on δ_a and $\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| \leq \delta_a$ ensure that

$$\sqrt{c_b\|\hat{\boldsymbol{a}}^{\langle t\rangle}\|} \leq \frac{1}{2}(1-\kappa_\theta)\sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|.$$

Hence with induction assumption that $\|\hat{\boldsymbol{\theta}}^{\langle t \rangle} - \boldsymbol{\theta}^\star\| \leq \sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|$, we have

$$\begin{aligned}
\|\hat{\boldsymbol{\theta}}^{\langle t+1 \rangle} - \boldsymbol{\theta}^\star\| &\leq \kappa_\theta \|\hat{\boldsymbol{\theta}}^{\langle t \rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_b \|\hat{\boldsymbol{a}}^{\langle t \rangle}\|} + \epsilon_\theta \\
&\leq \kappa_\theta \sqrt{1-(\kappa_a)^2}\|\theta\| \\
&\quad + \frac{1}{2}(1-\kappa_\theta)\sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\| + \frac{1}{2}(1-\kappa_\theta)\sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\| \\
&= \sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|.
\end{aligned}$$

Hence the second part of Equation (2.116) holds for $t+1$. This completes the proof.

□

2.6 Landscape of the Expected Log-likelihood

Do the results we derived in this chapter regarding the performance of EM provide any information on the landscape of our non-convex maximum likelihood optimization? To address this question, we show how our analysis can determine the stationary points of the expected log-likelihood and characterize the shape of the expected log-likelihood in a neighborhood of the stationary points. Let $L_f(\boldsymbol{\eta})$ denote the expected log-likelihood, i.e.,

$$L_f(\boldsymbol{\eta}) \triangleq \mathbb{E}(\log f_{\boldsymbol{\eta}}(\boldsymbol{Y})) = \int f(\boldsymbol{y};\boldsymbol{\eta}^\star) \log f(\boldsymbol{y};\boldsymbol{\eta}) \mathrm{d}\boldsymbol{y},$$

where $\boldsymbol{\eta}^\star$ denotes the true parameter value. Also consider the following standard regularity conditions:

R1 The family of probability density functions $f(\boldsymbol{y};\boldsymbol{\eta})$ have common support.

R2 $\nabla_{\boldsymbol{\eta}} \int f(\boldsymbol{y};\boldsymbol{\eta}^\star) \log f(\boldsymbol{y};\boldsymbol{\eta}) \mathrm{d}\boldsymbol{y} = \int f(\boldsymbol{y};\boldsymbol{\eta}^\star) \nabla_{\boldsymbol{\eta}} \log f(\boldsymbol{y};\boldsymbol{\eta}) \mathrm{d}\boldsymbol{y}$, where $\nabla_{\boldsymbol{\eta}}$ denotes the gradient with respect to $\boldsymbol{\eta}$.

R3 $\nabla_{\boldsymbol{\eta}}(\mathbb{E}\sum_{\boldsymbol{z}} f(\boldsymbol{z} \mid \boldsymbol{Y}; \boldsymbol{\eta}^{\langle t \rangle}) \log f(\boldsymbol{Y}, \boldsymbol{z}; \eta)) = \mathbb{E}\sum_{\boldsymbol{z}} f(\boldsymbol{z} \mid \boldsymbol{Y}; \boldsymbol{\eta}^{\langle t \rangle}) \nabla_{\boldsymbol{\eta}} \log f(\boldsymbol{Y}, \boldsymbol{z}; \boldsymbol{\eta})$.

Then we have the following lemma that establishes the connection between fixed points of the algorithms and the stationary points of the optimization problem.

Lemma 2.22. Let $\bar{\boldsymbol{\eta}} \in \mathbb{R}^d$ denote a stationary point of $L_f(\boldsymbol{\eta})$. Also assume that $Q(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{\langle t \rangle})$ has a unique and finite stationary point in terms of $\boldsymbol{\eta}$ for every $\boldsymbol{\eta}^{\langle t \rangle}$, and this stationary point is its global maxima. Then, if the model satisfies conditions R1–R3, and the Population EM algorithm is initialized at $\bar{\boldsymbol{\eta}}$, it will stay at $\bar{\boldsymbol{\eta}}$. Conversely, any fixed point of Population EM is a stationary point of $L_f(\boldsymbol{\eta})$.

Proof. Let $\bar{\boldsymbol{\eta}}$ denote a stationary point of $L_f(\boldsymbol{\eta})$. We first prove that $\bar{\boldsymbol{\eta}}$ is a stationary point of $Q(\boldsymbol{\eta} \mid \bar{\boldsymbol{\eta}})$ (See definition in Equation (2.1)).

$$\begin{aligned}
\nabla_{\boldsymbol{\eta}} Q(\boldsymbol{\eta} \mid \bar{\boldsymbol{\eta}})|_{\boldsymbol{\eta}=\bar{\boldsymbol{\eta}}} &= \int \sum_{\boldsymbol{z}} f(\boldsymbol{z} \mid \boldsymbol{y}; \bar{\boldsymbol{\eta}}) \frac{\nabla_{\boldsymbol{\eta}} f(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\eta})|_{\boldsymbol{\eta}=\bar{\boldsymbol{\eta}}}}{f(\boldsymbol{y}, \boldsymbol{z}; \bar{\boldsymbol{\eta}})} f(\boldsymbol{y}; \boldsymbol{\eta}^\star) \mathrm{d}\boldsymbol{y} \\
&= \int \sum_{\boldsymbol{z}} \frac{\nabla_{\boldsymbol{\eta}} f(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\eta})|_{\boldsymbol{\eta}=\bar{\boldsymbol{\eta}}}}{f(\boldsymbol{y}; \bar{\boldsymbol{\eta}})} f(\boldsymbol{y}; \boldsymbol{\eta}^\star) \mathrm{d}\boldsymbol{y} \\
&= \int \frac{\nabla_{\boldsymbol{\eta}} f(\boldsymbol{y}, \boldsymbol{\eta})|_{\boldsymbol{\eta}=\bar{\boldsymbol{\eta}}}}{f(\boldsymbol{y}; \bar{\boldsymbol{\eta}})} f(\boldsymbol{y}; \boldsymbol{\eta}^\star) \mathrm{d}\boldsymbol{y} = \mathbf{0}\,,
\end{aligned}$$

where the last equality is using the fact that $\bar{\boldsymbol{\eta}}$ is a stationary point of $L_f(\boldsymbol{\eta})$. Since $Q(\boldsymbol{\eta} \mid \bar{\boldsymbol{\eta}})$ has a unique stationary point, and we have assumed that the unique stationary point is its global maxima, then Population EM will stay at that point. The proof of the other direction is similar. □

Remark 2.6. The fact that $\boldsymbol{\eta}^\star$ is the global maximizer of $L_f(\boldsymbol{\eta})$ is well-known in the statistics and machine learning literature [e.g., Conniffe, 1987]. Furthermore, the fact that $\boldsymbol{\eta}^\star$ is a global maximizer of $Q(\boldsymbol{\eta} \mid \boldsymbol{\eta}^\star)$ is known as the self-consistency property Balakrishnan *et al.* [2017].

Back to our models, it is clear that Model 1 and Model 2 satisfy the conditions R1-R3. Therefore, we have the following corollary that analyzes the landscape of Model 1 and Model 2.

Corollary 2.2.

- The expected log-likelihood objective for Model 1 has only three stationary points. If $d = 1$ (so $\boldsymbol{\theta} = \theta \in \mathbb{R}$), then 0 is a local minima, while $\theta^\star$ and $-\theta^\star$ are global maxima. If $d > 1$, then $\mathbf{0}$ is a saddle point, and $\boldsymbol{\theta}^\star$ and $-\boldsymbol{\theta}^\star$ are global maxima.

- The expected log-likelihood objective for Model 2 has only three stationary points:

$$\left(\frac{\boldsymbol{\mu}_1^\star + \boldsymbol{\mu}_2^\star}{2}, \frac{\boldsymbol{\mu}_2^\star - \boldsymbol{\mu}_1^\star}{2}\right), \quad \left(\frac{\boldsymbol{\mu}_1^\star + \boldsymbol{\mu}_2^\star}{2}, \frac{\boldsymbol{\mu}_1^\star - \boldsymbol{\mu}_2^\star}{2}\right) \quad \text{and} \quad \left(\frac{\boldsymbol{\mu}_1^\star + \boldsymbol{\mu}_2^\star}{2}, \frac{\boldsymbol{\mu}_1^\star + \boldsymbol{\mu}_2^\star}{2}\right).$$

  The first two points are global maxima. The third point is a local minima when $d = 1$ or a saddle point when $d > 1$.

Finally, we have the following theorem that analyze the landscape for Model 4.

Theorem 2.8. For all $w_1^\star \neq 0.5$, the expected log-likelihood objective for Model 4 has only one saddle point $(\boldsymbol{\theta}, w_1) = (\mathbf{0}, 1/2)$ and no local maximizers besides the two global maximizers $(\boldsymbol{\theta}, w_1) = (\boldsymbol{\theta}^\star, w_1^\star)$ and $(\boldsymbol{\theta}, w_1) = (-\boldsymbol{\theta}^\star, w_2^\star)$.

The proof of this theorem is slightly more complicated because Model 4 does not satisfy the conditions R1-R3 on the boundary when $w_1 = 0$ or $w_1 = 1$. Hence, we can not directly apply Lemma 2.22 and we leave the proof in Appendix A.2.2.1.

Remark 2.7. Consider the landscape of the expected log-likelihood objective for Model 3 and the point $(\theta_{\text{wrong}}, w_1^\star)$, where θ_{wrong} is the local maximizer suggested by Theorem 2.3. Theorem 2.8 implies that we can still easily escape this point due to the non-zero gradient in the direction of w_1 and thus $(\theta_{\text{wrong}}, w_1^\star)$ is not even a saddle

point. We emphasize that this is exactly the mechanism that we have hoped for the purpose and benefit of over-parameterization (See the left panel in Figure 2.2).

Remark 2.8. Note that although $(\boldsymbol{\theta}, w_1) = ((w_1^\star - w_2^\star)\boldsymbol{\theta}^\star, 1)$ or $((w_2^\star - w_1^\star)\boldsymbol{\theta}^\star, 0)$ are the two fixed points for Population EM with Model 4 as well, they are not the first order stationary points of the expected log-likelihood objective if $w_1^\star \neq 0.5$.

Chapter 3

Approximate Message Passing Framework for Phase Retrieval

We first formally introduce our message passing algorithm. Following the steps proposed in Rangan [2011], we obtain the following algorithm called, *Approximate Message Passing for Amplitude-based optimization* (AMP.A). Starting from an initial estimate $\boldsymbol{x}^0 \in \mathbb{C}^{n\times 1}$, AMP.A proceeds as follows for $t \geq 0$:

$$\begin{aligned}\boldsymbol{p}^t &= \boldsymbol{A}\boldsymbol{x}^t - \frac{\lambda_{t-1}}{\delta} \cdot \frac{g(\boldsymbol{p}^{t-1}, \boldsymbol{y})}{-\mathrm{div}_p(g_{t-1})},\\ \boldsymbol{x}^{t+1} &= \lambda_t \cdot \left(\boldsymbol{x}^t + \boldsymbol{A}^{\mathrm{H}} \frac{g(\boldsymbol{p}^t, \boldsymbol{y})}{-\mathrm{div}_p(g_t)}\right).\end{aligned}$$

In these iterations

$$g(p, y) = y \cdot \frac{p}{|p|} - p,$$

and

$$\lambda_t = \frac{-\mathrm{div}_p(g_t)}{-\mathrm{div}_p(g_t) + \mu_k\left(\tau_t + \frac{1}{2}\right)},$$
$$\tau^t = \frac{1}{\delta}\frac{\tau^{t-1} + \frac{1}{2}}{-\mathrm{div}_p(g_{t-1})} \cdot \lambda_{t-1}.$$

In the above, $p/|p|$ at $p = 0$ can be any fixed number and does not affect the performance of AMP.A. Further, the "divergence" term $\mathrm{div}_p(g_t)$ is defined as

$$\begin{aligned}\mathrm{div}_p(g_t) &\triangleq \frac{1}{m}\sum_{a=1}^{m}\frac{1}{2}\left(\frac{\partial g(p_a^t, y_a)}{\partial p_a^R} - \mathrm{i}\frac{\partial g(p_a^t, y_a)}{\partial p_a^I}\right) \\ &= \frac{1}{m}\sum_{a=1}^{m}\frac{y_a}{2|p_a^t|} - 1,\end{aligned} \tag{3.2}$$

where p_a^R and p_a^I denote the real and imaginary parts of p_a^t respectively (i.e., $p_a^t = p_a^R + \mathrm{i}p_a^I$).

Note that we add an regularization term to the amplitude loss and here we would like to discuss the effect of this regularizer on AMP.A. For the moment suppose that the noise $\boldsymbol{w}$ is zero. Does including the regularizer in Equation (1.4) benefit AMP.A? Clearly, any regularization may introduce unnecessary bias to the solution. Hence, if the final goal is to obtain $\boldsymbol{x}^\star$ exactly we should set $\mu_k = 0$. However, the optimization problem in Equation (1.4) is non-convex and iterative algorithms intended to solve it can get stuck at bad local minima. In this regard, regularization can still help AMP.A to escape bad local minima through continuation. Continuation is popular in convex optimization for improving the convergence rate of iterative algorithms [Hale *et al.*, 2008]. In continuation we start with a value of μ_k for which AMP.A is capable of finding the global minimizer of Equation (1.4). Then, once AMP.A converges we will either decrease or increase μ_k a little bit (depending on the final value of μ for which

we want to solve the problem) and use the previous fixed point of AMP.A as the initialization for the new AMP.A. We continue this process until we reach the value of μ_k we are interested in. For instance, if we would like to solve the noiseless phase retrieval problem then μ_k should eventually go to zero so that we do not introduce unnecessary bias. The rationale behind continuation is the following. Let μ_k and μ_k' be two different values of the regularization parameter, and they are close to each other. Suppose that the global minimizer of Equation (1.4) with regularization parameter μ_k' is $\boldsymbol{x}(\mu_k')$ and is given to the user. Suppose further that the user would like to find the global minimizer of Equation (1.4) with μ_k. Then, it is conceivable that the global minimizer of the new problem is close to $\boldsymbol{x}(\mu_k')$.[1] Hence, the user can initialize AMP.A with $\boldsymbol{x}(\mu_k')$ and hope that the algorithm may converge to the global minimizer of Equation (1.4) for μ_k.

A more general version of the continuation idea we discussed above is to let μ_k change at every iteration (denoted as μ_k^t), and set λ_t according to μ_k^t:

$$\lambda_t = \frac{-\mathrm{div}_p(g_t)}{-\mathrm{div}_p(g_t) + \mu_k^t \left(\tau_t + \frac{1}{2}\right)}, \tag{3.3}$$

This way we can not only automate the continuation process, but also let AMP.A decide which choice of μ_k is appropriate at a given stage of the algorithm. Our discussion so far has been heuristic. It is not clear whether and how much the generalized continuation can benefit the algorithm. To give a partial answer to this question we focus on the following particular continuation strategy: $\mu_k^t = \frac{1+2\mathrm{div}_p(g_t)}{1+2\tau_t}$ and obtain

[1] Given the sometimes complex geometry of non-convex problems, this might not always be the case.

the following version of AMP.A:

$$\boldsymbol{p}^t = \boldsymbol{A}\boldsymbol{x}^t - \frac{2}{\delta} g(\boldsymbol{p}^{t-1}, \boldsymbol{y}), \tag{3.4a}$$

$$\boldsymbol{x}^{t+1} = 2\left[-\mathrm{div}_p(g_t)\cdot \boldsymbol{x}^t + \boldsymbol{A}^{\mathrm{H}} g(\boldsymbol{p}^t, \boldsymbol{y})\right]. \tag{3.4b}$$

3.1 Asymptotic analysis of AMP.A

In this section, we present the asymptotic platform under which AMP.A is studied, and we derive a set of equations, known as state evolution (SE), that capture the performance of AMP.A under the asymptotic analysis. The results are combinations of our paper Ma *et al.* [2018] and Ma *et al.* [2019].

3.1.1 Asymptotic framework and state evolution

Our analysis of AMP.A is carried out based on a standard asymptotic framework developed in Bayati and Montanari [2011, 2012]. In this framework, we let $m, n \to \infty$, while $m/n \to \delta$. Within this section, we will write $\boldsymbol{x}^\star$, $\boldsymbol{x}^t$, $\boldsymbol{w}$ and $\boldsymbol{A}$ as $\boldsymbol{x}^\star(n)$, $\boldsymbol{x}^t(n)$, $\boldsymbol{w}(n)$ and $\boldsymbol{A}(n)$ to make explicit their dependency on the signal dimension n. In this section we focus on the complex-valued AMP. We postpone the discussion of the real-valued AMP until Section 3.2. Following Mousavi *et al.* [2015], we introduce the following definition of converging sequences.

Definition 2. The sequence of instances $\{\boldsymbol{x}^\star(n), \boldsymbol{A}(n), \boldsymbol{w}(n)\}$ is said to be a converging sequence if the following hold:

- $\frac{m}{n} \to \delta \in (0, \infty)$, as $n \to \infty$.
- $\boldsymbol{A}(n)$ has i.i.d. Gaussian entries where $A_{ij} \sim \mathcal{CN}(0, 1/m)$.

– The empirical distribution of $\boldsymbol{x}^\star(n) \in \mathbb{C}^n$ converges weakly to a probability measure p_X with bounded second moment. Further, $\frac{1}{n}\|\boldsymbol{x}^\star(n)\|^2 \to \kappa^2$ where $\kappa^2 \in (0,\infty)$ is the second moment of p_X. For convenience and without loss of generality, we assume $\kappa = 1$.[2]

– The empirical distribution of $\boldsymbol{w}(n) \in \mathbb{C}^n$ converges weakly to $\mathcal{CN}(0,\sigma_w^2)$.

Under the asymptotic framework introduced above, the behavior of AMP.A can be characterized exactly. Roughly speaking, the estimate produced by AMP.A in each iteration is approximately distributed as the (scaled) true signal + additive Gaussian noise; in other words, $\boldsymbol{x}^t$ can be modeled as $\alpha_t\boldsymbol{x}^\star + \sigma_t\boldsymbol{h}$, where $\boldsymbol{h}$ behaves like an iid standard complex normal noise. We will clarify this claim in Theorem 3.1 below. The scaling constant α_t and the noise standard deviation σ_t evolve according to a known deterministic rule, called the state evolution (SE), defined below.

Definition 3. Starting from fixed $(\alpha_0,\sigma_0^2) \in \mathbb{C}\times\mathbb{R}_+\backslash(0,0)$, the sequences $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ are generated via the following recursion:

$$\begin{aligned}\alpha_{t+1} &= \psi_1(\alpha_t,\sigma_t^2),\\ \sigma_{t+1}^2 &= \psi_2(\alpha_t,\sigma_t^2;\delta,\sigma_w^2),\end{aligned} \tag{3.5}$$

where $\psi_1 : \mathbb{C}\times\mathbb{R}_+ \mapsto \mathbb{C}$ and $\psi_2 : \mathbb{C}\times\mathbb{R}_+ \mapsto \mathbb{R}_+$ are respectively given by

$$\begin{aligned}\psi_1(\alpha,\sigma^2) &= 2\cdot\mathbb{E}\left[\partial_z g(P,Y)\right] = \mathbb{E}\left[\frac{\bar{Z}P}{|Z|\,|P|}\right],\\ \psi_2(\alpha,\sigma^2;\delta,\sigma_w^2) &= 4\cdot\mathbb{E}\left[|g(P,Y)|^2\right] = 4\cdot\mathbb{E}\left[(|P|-|Z|-W)^2\right].\end{aligned}$$

[2]Otherwise, we can introduce the following normalized variables: $\tilde{\boldsymbol{y}} = \boldsymbol{y}/\kappa$, $\tilde{\boldsymbol{x}} = \boldsymbol{x}/\kappa$, $\tilde{\boldsymbol{w}} = \boldsymbol{w}/\kappa$, $\tilde{\boldsymbol{x}}^t = \boldsymbol{x}^t/\kappa$ and $\tilde{\boldsymbol{p}}^t = \boldsymbol{p}^t/\kappa$. One can verify that the AMP.A algorithm defined in Equation (3.4) for these normalized variables remains unchanged. Therefore, we can view that our analyses are carried out for these normalized variables; we don't need to actually change the algorithm though.

In the above equations, the expectations are over all random variables involved: $Z \sim \mathcal{CN}(0, 1/\delta)$, $P = \alpha Z + \sigma B$ where $B \sim \mathcal{CN}(0, 1/\delta)$ is independent of Z, and $Y = |Z| + W$ where $W \sim \mathcal{CN}(0, \sigma_w^2)$ is independent of both Z and B. Further, the partial Wirtinger derivative $\partial_z g(p, |z| + w)$ is defined as:

$$\partial_z g(p, |z| + w) \triangleq \frac{1}{2}\left[\frac{\partial}{\partial z_R} g(p, |z| + w) - \mathrm{i}\frac{\partial}{\partial z_I} g(p, |z| + w)\right],$$

where z_R and z_I are the real and imaginary parts of z (i.e., $z = z_R + \mathrm{i}z_I$).

Remark 3.1. The functions ψ_1 and ψ_2 are well defined except when both α and σ^2 are zero.

Remark 3.2. Most of the analysis in this chapter is concerned with the noiseless case. For brevity, we will often write $\psi_2(\alpha, \sigma; \delta, 0)$ (where $\sigma_w^2 = 0$) as $\psi_2(\alpha, \sigma; \delta)$. Further, when our focus is on α and σ^2 rather than δ, we will simply write $\psi_2(\alpha, \sigma^2; \delta)$ as $\psi_2(\alpha, \sigma^2)$.

In Appendix A.3.1.2, we simplify the functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ into the following expressions (with θ_α being the phase of α):

$$\psi_1(\alpha, \sigma^2) = e^{\mathrm{i}\theta_\alpha} \cdot \int_0^{\frac{\pi}{2}} \frac{|\alpha| \sin^2\theta}{\left(|\alpha|^2 \sin^2\theta + \sigma^2\right)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{3.6a}$$

$$\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) = \frac{4}{\delta}\left(|\alpha|^2 + \sigma^2 + 1 - \int_0^{\frac{\pi}{2}} \frac{2|\alpha|^2 \sin^2\theta + \sigma^2}{\left(|\alpha|^2 \sin^2\theta + \sigma^2\right)^{\frac{1}{2}}} \mathrm{d}\theta\right) + 4\sigma_w^2. \tag{3.6b}$$

The above expressions for ψ_1 and ψ_2 are more convenient for our analysis.

The state evolution framework for generalized AMP (GAMP) algorithms [Rangan, 2011] was first introduced and analyzed in Rangan [2011] and later formally proved in Javanmard and Montanari [2013]. As we will show later in Theorem 3.1, SE characterizes the macroscopic behavior of AMP.A. To apply the results in Rangan [2011];

Javanmard and Montanari [2013] to AMP.A, however, we need two generalizations. First, we need to extend the results to complex-valued models. This is straightforward by applying a complex-valued version of the conditioning lemma introduced in Rangan [2011]; Javanmard and Montanari [2013]. Second, existing results in Rangan [2011]; Javanmard and Montanari [2013] require the function g to be smooth. Our simulation results in case of complex-valued AMP.A show that SE predicts the performance of AMP.A despite the fact that g is not smooth. Since the thesis mainly focus on the convergence of the algorithm, we use the smoothing idea discussed in Zheng *et al.* [2017] to address the issues and connect the SE equations presented in Equation (3.5) with the iterations of AMP.A in Equation (3.4). Let $\epsilon > 0$ be a small fixed number. Consider the following smoothed version of AMP.A:

$$\begin{aligned}\boldsymbol{p}^t &= \boldsymbol{A}\boldsymbol{x}_{\boldsymbol{\epsilon}}{}^t - \frac{2}{\delta} g_\epsilon(\boldsymbol{p}^{t-1}, \boldsymbol{y}),\\ \boldsymbol{x}_{\boldsymbol{\epsilon}}{}^{t+1} &= 2\left[-\mathrm{div}_p(g_{t,\epsilon}) \cdot \boldsymbol{x}_{\boldsymbol{\epsilon}}{}^t + \boldsymbol{A}^{\mathrm{H}} g_\epsilon(\boldsymbol{p}^t, \boldsymbol{y})\right],\end{aligned}$$

where $g_\epsilon(\boldsymbol{p}^{t-1}, \boldsymbol{y})$ refers to a vector produced by applying $g_\epsilon : \mathbb{C} \times \mathbb{R}_+ \mapsto \mathbb{C}$ below component-wise:

$$g_\epsilon(p, y) \triangleq y \cdot h_\epsilon(p) - p,$$

where for $p = p_1 + \mathrm{i}p_2$, $h_\epsilon(p)$ is defined as

$$h_\epsilon(p) \triangleq \frac{p_1 + \mathrm{i}p_2}{\sqrt{p_1^2 + p_2^2 + \epsilon}}.$$

Note that as $\epsilon \to 0$, $g_{t,\epsilon} \to g_t$ and hence we expect the iterations of smoothed-AMP.A converge to the iterations of AMP.A.

Theorem 3.1 (asymptotic characterization). Let $\{\boldsymbol{x}^\star(n), \boldsymbol{A}(n), \boldsymbol{w}(n)\}$ be a converging sequence of instances. For each instance, let $\boldsymbol{x}^0(n)$ be an initial estimate independent

of $\boldsymbol{A}(n)$. Assume that the following hold almost surely

$$\lim_{n\to\infty} \frac{1}{n}\langle \boldsymbol{x}^\star, \boldsymbol{x}^0\rangle = \alpha_0 \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x}^0\|^2 = \sigma_0^2 + |\alpha_0|^2.$$

Let $\boldsymbol{x}_\epsilon^t(n)$ be the estimate produced by the smoothed AMP.A initialized by $\boldsymbol{x}^0(n)$ (which is independent of $\boldsymbol{A}(n)$) and $\boldsymbol{p}^{-1}(n) = \boldsymbol{0}$. Let $\epsilon_1, \epsilon_2, \ldots$ denote a sequence of smoothing parameters for which $\epsilon_i \to 0$ as $i \to \infty$ Then, for any iteration $t \geq 1$, the following holds almost surely

$$\lim_{j\to\infty} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |x_{\epsilon_j,i}^t(n) - e^{\mathrm{i}\theta_t}\, x_i^\star|^2 = \mathbb{E}\left[|X^t - e^{\mathrm{i}\theta_t} X^\star|^2\right] = \left|1 - |\alpha_t|\right|^2 + \sigma_t^2, \qquad (3.8)$$

where $\theta_t = \angle \alpha_t$, $X^t = \alpha_t X^\star + \sigma_t H$ and $X^\star \sim p_X$ is independent of $H \sim \mathcal{CN}(0,1)$. Further, $\{\alpha\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ are determined by Equation (3.5) with initialization α_0 and σ_0^2.

The proof of Theorem 3.1 is given in Appendix A.3.2.

3.1.2 Convergence of the SE for noiseless model

We now analyze the dynamical behavior of the SE. Before we proceed, we point out that in phase retrieval, one can only hope to recover the signal up to global phase ambiguity [Netrapalli *et al.*, 2013; Candès *et al.*, 2013, 2015], for generic signals without any structure. In light of Equation (3.8), AMP.A is successful if $|\alpha_t| \to 1$ and $\sigma_0^2 \to 0$ as $t \to \infty$.

Let us start with the following interesting feature of the state evolution, which can be seen from Equation (3.6).

Lemma 3.1. For any $(\alpha_0, \sigma_0^2) \in \mathbb{C} \times \mathbb{R}_+ \backslash (0,0)$, ψ_1 and ψ_2 satisfy the following properties:

(i) $\psi_1(\alpha, \sigma^2) = \psi_1(|\alpha|, \sigma^2) \cdot e^{\mathrm{i}\theta_\alpha}$, with $e^{\mathrm{i}\theta_\alpha}$ being the phase of α;

(ii) $\psi_2(\alpha, \sigma^2) = \psi_2(|\alpha|, \sigma^2)$.

Hence, if θ_t denotes the phase of α_t, then $\theta_t = \theta_0$.

In light of this lemma, we can focus on real and nonnegative values of α_t. In particular, we assume that $\alpha_0 \geq 0$ and we are interested in whether and under what conditions can the SE converge to the fixed point $(\alpha, \sigma^2) = (1, 0)$. The following two values of δ will play critical roles in the analysis of SE:

$$\delta_{\mathrm{AMP}} \triangleq \frac{64}{\pi^2} - 4 \approx 2.5,$$
$$\delta_{\mathrm{global}} \triangleq 2.$$

Our next theorem reveals the importance of δ_{AMP}. The proof of this theorem detailed in Section 3.3.

Theorem 3.2 (convergence of SE). Consider the noiseless model where $\sigma_w^2 = 0$. If $\delta > \delta_{\mathrm{AMP}}$, then for any $0 < |\alpha_0| \leq 1$ and $\sigma_0^2 \leq 1$, the sequences $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ defined in Equation (3.5) converge to

$$\lim_{t\to\infty} |\alpha_t| = 1 \quad \text{and} \quad \lim_{t\to\infty} \sigma_t^2 = 0.$$

There are a couple of points that we would like to emphasize here:

1. $0 < |\alpha_0| \leq 1$ and $\sigma_0^2 \leq 1$ is a pessimistic condition for Theorem 3.2. In particular, when $\delta > 4$, this condition could be relaxed to $\alpha_0 \neq 0$ and $\sigma_0^2 < \infty$. In this chapter, we did not try to optimize this condition since it is fairly loose and can be achieved by the spectral method in the noiseless case. In other words, if $\delta > \delta_{\mathrm{AMP}}$ the issue of initialization becomes minor. Alternatively,

$0 < |\alpha_0| \leq 1$ and $\sigma_0^2 \leq 1$ can also be achieved for the noiseless setting if the signal of interest has nonzero mean. To see this, consider the initialization $\boldsymbol{x}^0 = \mathbf{1}$. (In the general case where $\kappa \neq 1$, we initialize as $\boldsymbol{x}^0 = \kappa\mathbf{1}$. Note that $\kappa^2 = \|\boldsymbol{x}\|^2/n$ can be accurately estimated in the noiseless setting [Lu and Li, 2017].) Such initialization ensures that $|\alpha_0|^2+\sigma_0^2 = 1$. Further, $\alpha_0 = \mathbb{E}[X^\star] \neq 0$. Therefore, $|\alpha_0| \in (0,1)$ and $\sigma_0^2 \in (0,1)$.

2. $\alpha_0 \neq 0$ is essential for the success of AMP.A. This can be seen from the fact that $\alpha = 0$ is always a fixed point of $\psi_1(\alpha, \sigma^2)$ for any $\sigma^2 > 0$. From our definition of α_0 in Theorem 3.1, $\alpha_0 = 0$ is equivalent to $\frac{1}{n}\langle \boldsymbol{x}^\star, \boldsymbol{x}^0\rangle = 0$. This means that the initial estimate $\boldsymbol{x}^0$ cannot be orthogonal to the true signal vector $\boldsymbol{x}^\star$, otherwise there is no hope to recover the signal no matter how large δ is.

The following theorem describes the importance of δ_{global} and its proof can be found in Section 3.4.

Theorem 3.3 (local convergence of SE). When $\sigma_w^2 = 0$, then $(\alpha, \sigma^2) = (1, 0)$ is a fixed point of the SE in Equation (3.6). Furthermore, if $\delta > \delta_{\text{global}}$, then there exist two constants $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that the SE converges to this fixed point for any $\alpha_0 \in (1-\epsilon_1, 1)$ and $\sigma_0^2 \in (0, \epsilon_2)$. On the other hand if $\delta < \delta_{\text{global}}$, then the SE cannot converge to $(1, 0)$ except when initialized there.

According to Theorem 3.3, with proper initialization, SE can potentially converge to (α, σ^2) even if $\delta_{\text{global}} < \delta < \delta_{\text{AMP}}$. However, there are two points we should emphasize here: (i) As δ decreases from δ_{AMP} to δ_{global} the basin of attraction of $(\alpha, \sigma^2) = (1, 0)$ shrinks. Check the numerical results in Figure 3.1. (ii) we find that when $\delta < \delta_{\text{AMP}}$, standard initialization techniques, such as the spectral method, do not help AMP.A converge to $\boldsymbol{x}^\star$. More specifically, to find out whether spectral initialization helps our algorithm, we need to examine whether (α_0, σ_0^2) produced by

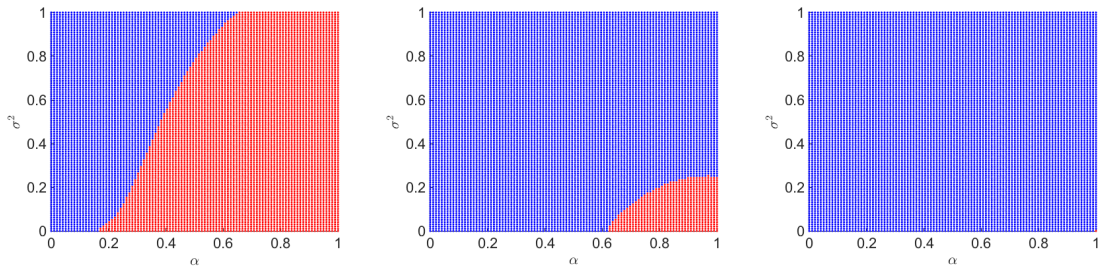

Figure 3.1: The red region exhibits the basin of attraction of $(\alpha, \sigma^2) = (1, 0)$. From left to right $\delta = 2.45$, $\delta = 2.3$, $\delta = 2.1$. Note that the basin of attraction of $(1, 0)$ in the case of $\delta = 2.1$ is a really small region in the bottom-right corner of the graph. The results are obtained by running the state evolution (SE) of AMP.A (complex-valued version) with α_0 and σ_0^2 chosen from 100×100 values equispaced in $[0, 1] \times [0, 1]$.

the spectral estimate can fall into the attraction basin of the good fixed point $(\alpha, \sigma^2) = (1, 0)$. Currently, the basin of attraction cannot be analytically characterized, but it can be conveniently computed via SE. Specifically, for a given (α_0, σ_0^2), we run the SE for a sufficiently large number of iterations and see if it converges to $(1, 0)$ (up to a pre-defined tolerance). On the other hand, the spectral initialization method was introduced in Netrapalli *et al.* [2013] for phase retrieval and subsequently studied in Candès *et al.* [2015]; Chen and Candès [2017]; Wang *et al.* [2016]; Lu and Li [2017]; Mondelli and Montanari [2017]. Specifically, the "direction" of the signal is estimated by the principal eigenvector $\boldsymbol{v}$ ($\|\boldsymbol{v}\|^2 = n$) of the following matrix:

$$\boldsymbol{D} \triangleq \boldsymbol{A}^{\mathrm{H}} \mathrm{diag}\{\mathcal{T}(y_1), \ldots, \mathcal{T}(y_m)\} \boldsymbol{A}, \tag{3.9}$$

where $\mathcal{T} : \mathbb{R}_+ \to \mathbb{R}$ is a nonlinear processing function, and $\mathrm{diag}\{a_1, \ldots, a_m\}$ is a diagonal matrix with diagonal entries given by $\{a_1, \ldots, a_m\}$. The exact asymptotic performance of the spectral method was characterized in Lu and Li [2017] under some regularity assumptions on $\mathcal{T}$.

Fig. 3.2 plots the basin of attraction of the fixed point $(\alpha, \sigma) = (1, 0)$ for $\delta = 2.4$

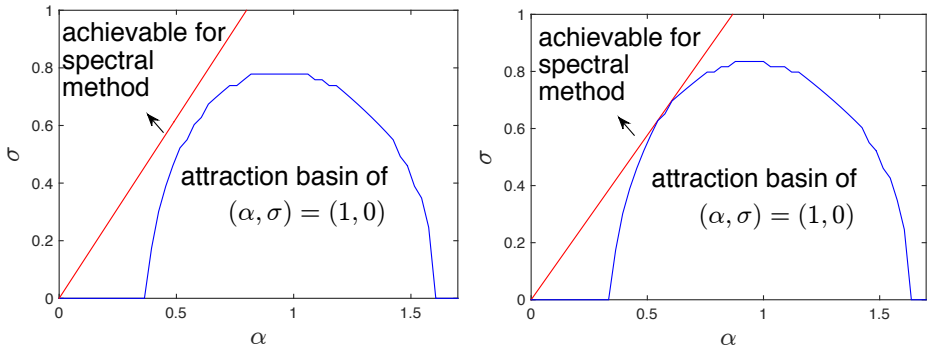

Figure 3.2: Plot of the attraction basin of AMP.A and the achievable region of the spectral method. **Left:** $\delta = 2.40$. **Right:** $\delta = 2.41$. In this figure, the vertical axis is σ instead of σ^2.

or 2.41 (indicated by the blue curve). The straight line is obtained in the following way: From Lu and Li [2017], for a given δ and $\mathcal{T}$, the ratio σ_0/α_0 can be computed by solving a set of fixed point equations, and this ratio determines a straight line $\sigma/\alpha = \sigma_0/\alpha_0$ in the $\alpha - \sigma$ plane. The red line in Fig. 3.2 is obtained using $\mathcal{T}$ derived in Mondelli and Montanari [2017]. The region above the red line can be potentially achieved by certain choices of $\mathcal{T}$ together with linear scaling. On the other hand, no known $\mathcal{T}$ can achieve the region below the red line. As we see in this figure, the spectral estimate cannot fall into the basin of attraction in the current example for $\delta = 2.4$ (left subfigure). The smallest δ such that two curves intersect is numerically found to be around $\delta = 2.41$ (right subfigure) which is quite close to $\delta_{\text{AMP}} \approx 2.48$. Notice that for $\delta > \delta_{\text{AMP}}$, AMP.A works for any $\alpha_0 \neq 0$. This means that the spectral method cannot help AMP.A much besides providing an estimate not orthogonal to the true signal. Hence, the question of finding initialization in the basin of attraction of $(\alpha, \sigma^2) = (1, 0)$ (when $\delta < \delta_{\text{AMP}}$) remains open for future research.

3.1.3 Noise sensitivity

So far we have only discussed the performance of AMP.A in the ideal setting where the noise is not present in the measurements. Now let us discuss the performance of AMP.A under noisy settings. We assume that the measurement noise is Gaussian and small. Clearly, in this setting exact recovery is impossible, hence we study the asymptotic mean square error defined as the following almost sure limit ($\theta_t \triangleq \angle \frac{1}{n}\langle \boldsymbol{x}^\star, \boldsymbol{x}^t \rangle$)

$$\mathrm{AMSE}(\sigma_w^2, \delta) \triangleq \lim_{t\to\infty} \frac{\|\boldsymbol{x}^t - e^{\mathrm{i}\theta_t}\boldsymbol{x}^\star\|_2^2}{n}, \tag{3.10}$$

In the noisy case, one can use Equation (3.5) to calculate the asymptotic MSE (AMSE) of AMP.A as a function of the variance of the noise and δ. However, as our next theorem demonstrates it is possible to obtain an explicit and informative expression for AMSE of AMP.A in the high signal-to-noise ratio (SNR) regime.

Theorem 3.4 (noise sensitivity). Suppose that $\delta > \delta_{\mathrm{AMP}} = \frac{64}{\pi^2} - 4$ and $0 < |\alpha_0| \leq 1$ and $\sigma_0^2 < 1$. Then, in the high SNR regime the asymptotic MSE defined in Equation (3.10) behaves as

$$\lim_{\sigma_w^2 \to 0} \frac{\mathrm{AMSE}(\sigma_w^2, \delta)}{\sigma_w^2} = \frac{4}{1 - \frac{2}{\delta}}.$$

The proof of this theorem can be found in Section 3.5.

3.1.4 Background on Elliptic Integrals

The functions that we have in Equation (3.5) are related to the first and second kinds of elliptic integrals. Below we review some of the properties of these functions that will be used throughout this chapter. Elliptic integrals (elliptic integral of the second kind) were originally proposed for the study of the arc length of ellipsoids. Since their appearance, elliptic integrals have appeared in many problems in physics

and chemistry, such as characterization of planetary orbits. Three types of elliptic integrals are of particular importance, since a large class of elliptic integrals can be reduced to these three. We introduce two of them that are of particular interest in our work.

Definition 4. The first and second kinds of complete elliptic integrals, denoted by $K(m)$ and $E(m)$ (for $-\infty < m < 1$) respectively, are defined as [Byrd and Friedman]

$$K(m) = \int_0^{\frac{\pi}{2}} \frac{1}{(1 - m\sin^2\theta)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{3.11a}$$

$$E(m) = \int_0^{\frac{\pi}{2}} (1 - m\sin^2\theta)^{\frac{1}{2}} \mathrm{d}\theta. \tag{3.11b}$$

For convenience, we also introduce the following definition:

$$T(m) = E(m) - (1 - m)K(m). \tag{3.11c}$$

In the above definitions, we continued to use m, to follow the convention in the literature of elliptic integrals. Previously, m was defined to be the number of measurements, but such abuse of notation should not cause confusion as the exact meaning of m is usually clear from the context.

Below, we list some properties of elliptic integrals that will be used in this chapter. The proofs of these properties can be found in standard references for elliptic integrals and thus omitted (e.g., Byrd and Friedman).

Lemma 3.2. The following hold for $K(m)$ and $E(m)$ defined in Equation (3.11):

(i) $K(0) = E(0) = \frac{\pi}{2}$. Further, for $\epsilon \to 0$, $E(1-\epsilon)$ and $K(1-\epsilon)$ behave as

$$\begin{aligned} E(1-\epsilon) &= 1 + \frac{\epsilon}{2}\left(\log\frac{4}{\sqrt{\epsilon}} - 0.5\right) + O(\epsilon^2 \log(1/\epsilon)) \\ K(1-\epsilon) &= \log\left(\frac{4}{\sqrt{\epsilon}}\right) + O(\epsilon \log(1/\epsilon)). \end{aligned}$$

(ii) On $m \in (0,1)$, $K(m)$ is strictly increasing, $E(m)$ is strictly decreasing, and $T(m)$ is strictly increasing.

(iii) For $m > -1$,

$$\begin{aligned} K(-m) &= \frac{1}{\sqrt{1+m}} K\left(\frac{m}{1+m}\right), \\ E(-m) &= \sqrt{1+m} E\left(\frac{m}{1+m}\right). \end{aligned}$$

(iv) The derivatives of $K(m)$, $E(m)$ and $T(m)$ are given by (for $m < 1$)

$$\begin{aligned} K'(m) &= \frac{E(m) - (1-m)K(m)}{2m(1-m)}, \\ E'(m) &= \frac{E(m) - K(m)}{2m}, \\ T'(m) &= \frac{1}{2}K(m). \end{aligned} \tag{3.13}$$

Furthermore, we will use a few more elliptic integrals in our work. Next lemma connects these elliptic integrals to Type I and Type II elliptic integrals.

Lemma 3.3. The following equalities hold for any $m \geq 0$:

$$\int_0^{\frac{\pi}{2}} \frac{\cos^2\theta}{\left(1+m\sin^2\theta\right)^{\frac{3}{2}}} \mathrm{d}\theta = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(1+m\sin^2\theta\right)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{3.14a}$$

$$\int_0^{\frac{\pi}{2}} \frac{3m\cos^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}} \mathrm{d}\theta + \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}} \mathrm{d}\theta = \int_0^{\frac{\pi}{2}} \frac{1+2m\sin^2\theta}{\left(1+m\sin^2\theta\right)^{\frac{1}{2}}} \mathrm{d}\theta. \tag{3.14b}$$

We prove this lemma in Appendix A.3.3

3.2 Extension to real-valued signals

Until now our focus is on complex-valued signals. In this section, our goal is to extend our results to real-valued signals. Since most of the results are similar to the complex-valued case, we will skip the details and only emphasize on the main differences.

3.2.1 AMP.A Algorithm

In the real-valued case, AMP.A uses the following iterations:

$$\begin{aligned}\boldsymbol{x}^{t+1} &= -\mathrm{div}_p(g_t)\cdot\boldsymbol{x}^t + \boldsymbol{A}^{\mathrm{T}} g(\boldsymbol{p}^t, \boldsymbol{y}),\\ \boldsymbol{p}^t &= \boldsymbol{A}\boldsymbol{x}^t - \frac{1}{\delta} g(\boldsymbol{p}^{t-1}, \boldsymbol{y}),\end{aligned}$$

where $g(p, y) : \mathbb{R}\times\mathbb{R}_+ \mapsto \mathbb{R}$ is given by

$$g(p, y) \triangleq y\cdot\mathrm{sign}(p) - p,$$

where $\mathrm{sign}(p)$ denotes the sign of p. We emphasize that the divergence term $\mathrm{div}_p(g_t)$ contains a Dirac delta at 0 due to the discontinuity of the sign function. This makes the calculation of the divergence in the AMP.A algorithm tricky. One can use the smoothing idea we discussed in Section 3.1.1. Alternatively, there are several possible approaches to estimate the divergence term.

3.2.2 Asymptotic Analysis

Our analysis is based on the same asymptotic framework detailed in Section 3.1.2. The only difference is that the measurement matrix is now real Gaussian with $A_{ij} \sim \mathcal{N}(0, 1/m)$ and $w_a \sim \mathcal{N}(0, \sigma_w^2)$. In the real-valued setting, the state evolution (SE) recursion of AMP.A in Equation (3.15) becomes the following.

Definition 5. Starting from fixed $(\alpha_0, \sigma_0^2) \in \mathbb{R} \times \mathbb{R}_+ \backslash (0,0)$ the sequences $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ are generated via the following iterations:

$$\begin{aligned} \alpha_{t+1} &= \psi_1(\alpha_t, \sigma_t^2), \\ \sigma_{t+1}^2 &= \psi_2(\alpha_t, \sigma_t^2; \delta, \sigma_w^2), \end{aligned} \tag{3.16}$$

where, with some abuse of notations, $\psi_1 : \mathbb{R} \times \mathbb{R}_+ \mapsto \mathbb{R}$ and $\psi_2 : \mathbb{R} \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ are now defined as

$$\begin{aligned} \psi_1(\alpha, \sigma^2) &= \mathbb{E}[\partial_z g(P, |Y|)] = \mathbb{E}[\text{sign}(Z\, P)], \\ \psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) &= \mathbb{E}[g^2(P, |Y|)] = \mathbb{E}\left[(|Z| - |P| + W)^2\right]. \end{aligned}$$

The expectations are over the following random variables: $Z \sim \mathcal{N}(0, 1/\delta)$, $P = \alpha Z + \sigma B$ where $B \sim \mathcal{N}(0, 1/\delta)$ is independent of Z, and $Y = |Z| + W$ where $W \sim \mathcal{N}(0, \sigma_w^2)$ independent of both Z and B.

In Appendix A.4.1, we derived the following closed-form expressions of ψ_1 and ψ_2:

$$\psi_1(\alpha, \sigma^2) = \frac{2}{\pi} \arctan\left(\frac{\alpha}{\sigma}\right), \tag{3.17a}$$

$$\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) = \frac{1}{\delta}\left[\alpha^2 + \sigma^2 + 1 - \frac{4\sigma}{\pi} - \frac{4\alpha}{\pi} \arctan\left(\frac{\alpha}{\sigma}\right)\right] + \sigma_w^2. \tag{3.17b}$$

As in the complex-valued case, we would like to study the dynamics of these two

equations. The following lemma simplifies the analysis.

Lemma 3.4. $\psi_1(\alpha, \sigma^2)$ and $\psi_2(\alpha, \sigma^2)$ in Equation (3.17) and Equation (3.17b) have the following properties:

(i) $\psi_1(\alpha, \sigma^2) = \psi_1(|\alpha|, \sigma^2) \cdot \text{sign}(\alpha)$.

(ii) $\psi_2(\alpha, \sigma^2) = \psi_2(|\alpha|, \sigma^2)$.

Again the following two values of δ play a critical role in the performance of AMP:

$$\delta_{\text{AMP}} = \frac{\pi^2}{4} - 1 \approx 1.47,$$
$$\delta_{\text{global}} = 1 + \frac{4}{\pi^2} \approx 1.40.$$

The following two theorems correspond to Theorems 3.2 and 3.3 that explain the dynamics of SE for complex-valued signals. The proofs can be found in Appendix A.4.2 and Appendix A.4.3 respectively.

Theorem 3.5 (convergence of SE). Suppose that $\delta > \delta_{\text{AMP}} = \frac{\pi^2}{4} - 1$ and $\sigma_w^2 = 0$. For any $\alpha_0 \in \mathbb{R}\backslash 0$ and $\sigma_0^2 < \infty$, the sequences $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ defined in Equation (3.16) converge:

$$\lim_{t\to\infty} |\alpha_t| = 1 \quad \text{and} \quad \lim_{t\to\infty} \sigma_t^2 = 0.$$

Note that in Theorem 3.5 the sequences converge for any $\sigma_0^2 < \infty$. This result is stronger than the complex-valued counterpart, which requires $0 < |\alpha_0| \leq 1$ and $\sigma_0^2 \leq 1$ (see Theorem 3.2).

Theorem 3.6 (local convergence of SE). For the noiseless setting where $\sigma_w^2 = 0$, $(\alpha, \sigma^2) = (1, 0)$ is a fixed point of the SE in Equation (3.6). Furthermore, if $\delta > \delta_{\text{global}} = 1 + \frac{4}{\pi^2}$, then there exist two constants $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that the SE converges to this fixed point for any $\alpha_0 \in (1 - \epsilon_1, 1)$ and $\sigma_0^2 \in (0, \epsilon_2)$. On the other hand if $\delta < \delta_{\text{global}}$, then the SE cannot converge to $(1, 0)$ except when initialized there.

Note that δ_{global} here is different from the information theoretic limit $\delta = 1$. We should emphasize that if we had not used the continuation discussed in Equation (3.3), then the basin of attraction of $(\alpha, \sigma) = (1, 0)$ would be non-empty as long as $\delta > 1$.

Finally, we discuss the performance of AMP.A in the high SNR regime. See Appendix A.4.4 for its proof.

Theorem 3.7 (noise sensitivity). Suppose that $\delta > \delta_{\text{AMP}} = \frac{\pi^2}{4} - 1$ and $\alpha_0 \in \mathbb{R}\backslash 0$ and $\sigma_0^2 < \infty$. Then, in the high SNR regime we have

$$\lim_{\sigma_w^2 \to 0} \frac{\text{AMSE}(\sigma_w^2, \delta)}{\sigma_w^2} = \frac{1}{\left(1 + \frac{4}{\pi^2}\right)^{-1} - \frac{1}{\delta}}.$$

3.3 Proof of Theorem 3.2

The goal of this section is to prove Theorem 3.2. However, since the proof is very long we start with the proof sketch to help the reader navigate through the complete proof.

3.3.1 Roadmap of the proof

Our main goal is to study the dynamics of the iterations:

$$\begin{aligned} \alpha_{t+1} &= \psi_1(\alpha_t, \sigma_t^2), \\ \sigma_{t+1}^2 &= \psi_2(\alpha_t, \sigma_t^2; \delta), \end{aligned} \tag{3.18}$$

Notice that according to the assumptions of the theorem, we assume that we initialized the dynamical system with $\alpha_0 > 0$. Our first hope is that this dynamical system will not oscillate and will converge to the solutions of the following system of nonlinear

equations:

$$\begin{aligned} \alpha &= \psi_1(\alpha, \sigma^2), \\ \sigma^2 &= \psi_2(\alpha, \sigma^2; \delta), \end{aligned} \tag{3.19}$$

Hence, the first step is to characterize and understand the fixed points of the solutions of Equation (3.19). Toward this goal we should study the properties of $\psi_1(\alpha, \sigma^2)$ and $\psi_2(\alpha, \sigma^2; \delta)$. In particular, we would like to know how the fixed points of $\psi_1(\alpha, \sigma^2)$ behave for a given σ^2 and how the fixed points of $\psi_2(\alpha, \sigma^2; \delta)$ behave for a given value of α and δ. The graphs of these functions are shown in Figure 3.3. We list some of

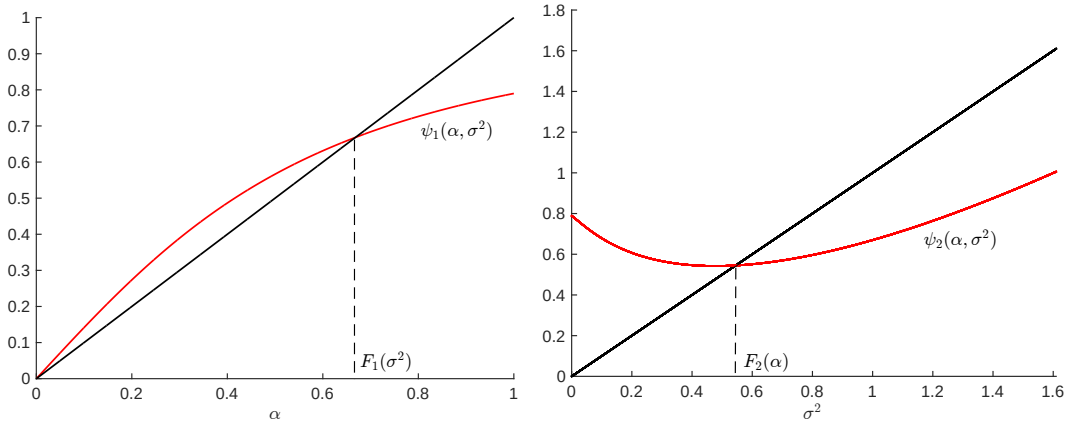

Figure 3.3: **Left:** plot of $\psi_1(\alpha, \sigma^2)$ against α. $\sigma^2 = 0.3$. **Right:** plot of $\psi_2(\alpha, \sigma^2; \delta)$ against σ^2. $\alpha = 0.3$ and $\delta = \delta_{\text{AMP}}$.

the important properties of these two functions. We refer the reader to Section 3.3.2 to see more accurate statement of these claims.

1. $\psi_1(\alpha, \sigma^2)$ is a concave and strictly increasing function of $\alpha > 0$, for any $\sigma^2 > 0$: This implies that $\psi_1(\alpha, \sigma^2)$ can have two fixed points: one at zero and one at $\alpha > 0$. Also, as is clear from the figure, the second fixed point is the stable one.

2. If $\delta > \delta_{\text{AMP}}$, then ψ_2 has always one stable fixed point. It may have one unstable

fixed points (as a function of σ^2). See Fig. A.1 for an example of this situation.

For the moment assume that the unstable fixed points do not affect the dynamics of AMP.A. Let $F_1(\sigma^2)$ denote the non-zero fixed point of ψ_1 and $F_2(\sigma^2)$ the stable fixed point of ψ_2.[3] We will prove in Lemma 3.11 that $F_1(\sigma^2)$ is a decreasing function and hence $F_1^{-1}(\alpha)$ is well-defined on $0 < \alpha \leq 1$. Moreover, we will show that by choosing $F_1^{-1}(0) = \frac{\pi^2}{16}$, $F_1^{-1}(\alpha)$ is continuous on [0,1]. $F_1^{-1}(\alpha)$ and $F_2(\alpha; \delta)$ are shown in Fig. 3.4. Note that the places these curves intersect correspond to the fixed points of Equation (3.19). Depending on the value of δ the two curves show the following different behaviors:

1. When $\delta > \delta_{\text{AMP}}$, the dashed curve (see Fig. 3.4) is entirely below the solid curve except at $(\alpha, \sigma^2) = (1, 0)$. δ_{AMP} is the critical value of δ at which $F_2(0; \delta) = F_1^{-1}(0)$. Formally, we will prove the following lemma:

   Lemma 3.5. If $\delta \geq \delta_{\text{AMP}} = \frac{64}{\pi^2} - 4$, then $F_1^{-1}(\alpha) > F_2(\alpha; \delta)$ holds for any $\alpha \in (0, 1)$.

   We prove this lemma in Appendix A.3.4.4.

   Intuitively speaking, in this case we expect the state evolution to converge to the fixed point $(\alpha, \sigma^2) = (1, 0)$, meaning that AMP.A achieves exact recovery.

2. When $2 < \delta < \delta_{\text{AMP}}$, the two curves intersect at multiple locations, but $F_2(\alpha) < F_1^{-1}(\alpha)$ for the values of α that are close to one. This implies that AMP.A can still exactly recover $\boldsymbol{x}^\star$ if the initialization is close enough to $\boldsymbol{x}^\star$. However, this does not happen with spectral initialization. We will discuss this case in Theorem 3.3 and we do not pursue it further here.

[3] In the literature of dynamical systems, these functions are sometimes called *nullclines*. Nullclines are useful for qualitatively analyzing local dynamical behavior of two-dimensional maps (which is the case for the SE in this chapter).

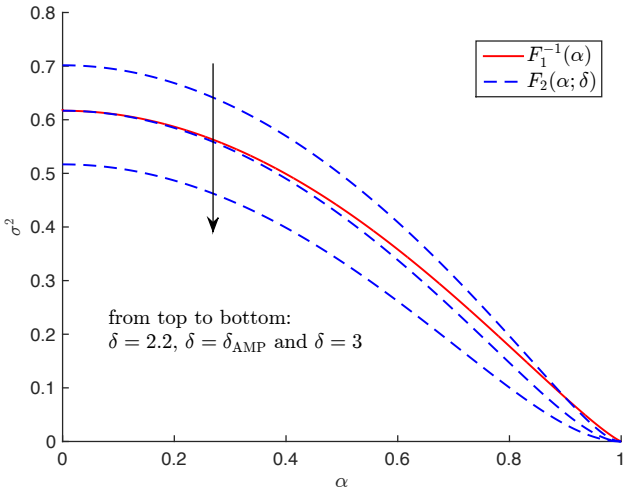

Figure 3.4: Plots of $F_1^{-1}(\alpha)$ and $F_2(\alpha)$ for different values of δ. When $\delta = \delta_{\text{AMP}}$, $F_1^{-1}(\alpha)$ and $F_2(\alpha;\delta)$ intersect at $\alpha = 0$.

So far, we have studied the solutions of Equation (3.19). But the ultimate goal of analysis of AMP.A is the analysis of Equation (3.18). In particular, it is important to show that the estimates (α_t, σ_t^2) converge to $(1, 0)$ and do not oscillate. Unfortunately, the dynamics of (α_t, σ_t^2) do not monotonically move toward the fixed point $(1, 0)$, which makes the analysis of SE complicated. More specifically, we can not directly apply Lemma 2.14 since $\psi_2(\alpha, \sigma^2; \delta)$ is not a monotone function of σ^2 (See Figure 3.3). Yet, since we can show the existence of the fixed point curves F_1 and F_2, it is possible to analyze the movements from (α_t, σ_t) to $(\alpha_{t+1}, \sigma_{t+1})$. We share some of the insights below.

Suppose that $\delta > \delta_{\text{AMP}}$. We first show that (α_t, σ_t^2) lies within a bounded region if the initialization falls into that region.

Lemma 3.6. Suppose that $\alpha_0 > 0$ and $\sigma_0^2 \leq 1$. If $\delta > \delta_{\text{AMP}} = \frac{64}{\pi^2} - 4$, then the sequences $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ generated by Equation (3.5) satisfy the following:

$$0 \leq \alpha_t \leq 1 \quad \text{and} \quad 0 \leq \sigma_t^2 \leq \sigma_{\max}^2, \quad \forall t \geq 1,$$

where $\sigma^2_{\max} \triangleq \max\left\{1, \frac{4}{\delta}\right\}$.

Proof. As discussed in Lemma 3.1, the assumption $\alpha_0 > 0$ implies that $\alpha_t > 0$, $\forall t \geq 1$. Further, from the property that $0 < \psi_1(\alpha, \sigma^2) < 1$ for $\alpha > 0$ and $\sigma^2 > 0$ (see Lemma 3.9 (ii)), we readily have $0 \leq \alpha_t \leq 1$. Similarly, Lemma 3.10 (iii) shows that if $\delta > \delta_{\text{AMP}}$, $\alpha \in [0, 1]$ and $\sigma^2 \in [0, \sigma^2_{\max}]$, then $0 \leq \psi_2(\alpha, \sigma^2; \delta) \leq \sigma^2_{\max}$. By our assumption, we have $\sigma_0^2 \leq 1 \leq \sigma^2_{\max}$, and using induction we prove $0 \leq \sigma_t^2 \leq \sigma^2_{\max}$. □

From the above lemma, we see that to understand the dynamics of the SE, we only focus on the region $\mathcal{R} \triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, 0 < \sigma^2 \leq \sigma^2_{\max}\right\}$. Since the dynamic of AMP.A is complicated, we divide this region into smaller regions. See Figure 3.5 for an illustration.

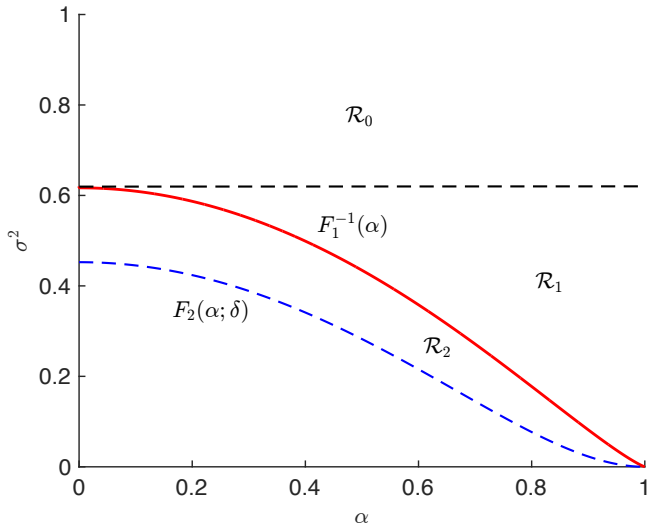

Figure 3.5: Illustration of the three regions in Definition 6. Note that $\mathcal{R}_2$ also includes the region below $F_2(\alpha; \delta)$.

Definition 6. We divide $\mathcal{R} \triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, 0 < \sigma^2 \leq \sigma^2_{\max}\right\}$ into the following

three sub-regions:

$$\begin{aligned}
\mathcal{R}_0 &\triangleq \left\{(\alpha,\sigma^2)\middle|0<\alpha\leq 1, \frac{\pi^2}{16}<\sigma^2\leq\sigma^2_{\max}\right\},\\
\mathcal{R}_1 &\triangleq \left\{(\alpha,\sigma^2)\middle|0<\alpha\leq 1, F_1^{-1}(\alpha)<\sigma^2\leq\frac{\pi^2}{16}\right\},\\
\mathcal{R}_2 &\triangleq \left\{(\alpha,\sigma^2)\middle|0<\alpha\leq 1, 0\leq\sigma^2\leq F_1^{-1}(\alpha)\right\}.
\end{aligned} \tag{3.20}$$

Our next lemma shows that if (α_t,σ_t^2) is in $\mathcal{R}_1$ or $\mathcal{R}_2$ for $t\geq 1$, then (α_t,σ_t^2) converges to $(1,0)$. The following lemma demonstrates this claim.

Lemma 3.7. Suppose that $\delta>\delta_{\text{AMP}}$. If $(\alpha_{t_0},\sigma_{t_0}^2)$ is in $\mathcal{R}_1\cup\mathcal{R}_2$ at time t_0 (where $t_0\geq 1$), and $\{\alpha_t\}_{t\geq t_0}$ and $\{\sigma_t^2\}_{t\geq t_0}$ are obtained via the SE in Equation (3.5), then

(i) (α_t,σ_t^2) remains in $\mathcal{R}_1\cup\mathcal{R}_2$ for all $t>t_0$;

(ii) (α_t,σ_t^2) converges:

$$\lim_{t\to\infty}\alpha_t=1 \quad \textit{and} \quad \lim_{t\to\infty}\sigma_t^2=0.$$

This claim will be proved in Section 3.3.3. Notice that the condition $t_0\geq 1$ is important for part (i) to hold: if (α_0,σ_0^2) is close to the origin (and thus in $\mathcal{R}_2$), then (α_1,σ_1^2) can move to $\mathcal{R}_0$. However, this cannot happen when $t\geq 1$. In the proof given in Section 3.3.3, we showed that for any $(\alpha_0,\sigma_0^2)\in\mathcal{R}$ the possible locations of (α_1,σ_1^2) are bounded from below by a curve, and once (α,σ^2) is above this curve and also in region $\mathcal{R}_1$ or $\mathcal{R}_2$, then we will prove that it cannot go to $\mathcal{R}_0$. Finally, we will prove the following Lemma that completes the proof.

Lemma 3.8. Suppose that $\delta>\delta_{\text{AMP}}$. Let $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ be the sequences generated according to Equation (3.5) from any $(\alpha_0,\sigma_0^2)\in\mathcal{R}_0$. Then, there exists a finite number $T\geq 1$ such that $(\alpha_T,\sigma_T^2)\in\mathcal{R}_1\cup\mathcal{R}_2$.

The proof of this result is in Section 3.3.4. Combining the above two lemmas, it is straightforward to see that $(\alpha_t, \sigma_t^2) \to (1, 0)$, and hence the proof is complete.

Below we present the missing details.

3.3.2 Properties of ψ_1, ψ_2, F_1 and F_2

First we present the main properties of ψ_1, ψ_2 and the stable fixed points functions F_1 and F_2 introduced in Section 3.3.1 that are useful throughout the chapter. The following first lemma summarizes the properties of ψ_1.

Lemma 3.9. $\psi_1(\alpha, \sigma^2)$ has the following properties (for $\alpha \geq 0$):

(i) $\psi_1(\alpha, \sigma^2)$ is a concave and strictly increasing function of $\alpha > 0$, for any given $\sigma^2 > 0$.

(ii) $0 < \psi_1(\alpha, \sigma^2) \leq 1$, for $\alpha > 0$ and $\sigma^2 > 0$.

(iii) If $0 < \sigma^2 < \pi^2/16$, then there are two nonnegative solutions to $\alpha = \psi_1(\alpha, \sigma^2)$: $\alpha = 0$ and $\alpha = F_1(\sigma^2) > 0$. Further, $F_1(\sigma^2)$ is strongly globally attracting, meaning that

$$\alpha < \psi_1(\alpha, \sigma^2) < F_1(\sigma^2), \quad \alpha \in (0, F_1(\sigma^2)), \tag{3.21a}$$

and

$$F_1(\sigma^2) < \psi_1(\alpha, \sigma^2) < \alpha, \quad \alpha \in (F_1(\sigma^2), \infty). \tag{3.21b}$$

On the other hand, if $\sigma^2 \geq \pi^2/16$ then $\alpha = 0$ is the unique nonnegative fixed point and it is strongly globally attracting.

We prove this lemma in Appendix A.3.4.1

Next, we present the properties of ψ_2 in the following lemma.

Lemma 3.10. $\psi_2(\alpha, \sigma^2; \delta)$ has the following properties:

(i) If $\delta < 2$, then $\sigma^2 = 0$ is a locally unstable fixed point to $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta)$, meaning that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}\Big|_{\alpha=1, \sigma^2=0} > 1.$$

(ii) For any $\delta > 2$, $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta)$ has a unique fixed point in $\sigma^2 \in [0, 1]$ for any $\alpha \in [0, 1]$. Further, the fixed point is (weakly) globally attracting in $\sigma^2 \in [0, 1]$:

$$\sigma^2 < \psi_2(\alpha, \sigma^2; \delta), \quad \sigma^2 \in (0, F_2(\alpha)), \tag{3.22a}$$

and

$$\psi_2(\alpha, \sigma^2) < \sigma^2, \quad \sigma^2 \in (F_2(\alpha), 1). \tag{3.22b}$$

(iii) If $\delta \geq \delta_{\text{AMP}}$, then for any $\alpha \in [0, 1]$, we have

$$0 \leq \psi_2(\alpha, \sigma^2; \delta) \leq \sigma^2_{\max}, \quad \sigma^2 \in [0, \sigma^2_{\max}],$$

where $\sigma^2_{\max} \triangleq \max\{1, 4/\delta\}$.

(iv) If $\delta \geq \delta_{\text{AMP}}$, then for any $\alpha \in [0, 1]$, $F_2(\alpha)$ is the unique (weakly) globally attracting fixed point of $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta)$ in $\sigma^2 \in [0, \sigma^2_{\max}]$. Namely,

$$\sigma^2 < \psi_2(\alpha, \sigma^2; \delta), \quad \sigma^2 \in (0, F_2(\alpha)), \tag{3.23a}$$

and

$$\psi_2(\alpha, \sigma^2) < \sigma^2, \quad \sigma^2 \in (F_2(\alpha), \sigma^2_{\max}). \tag{3.23b}$$

(v) For any $\delta > 0$, $\psi_2(\alpha, \sigma^2; \delta)$ is an increasing function of $\sigma^2 > 0$ if

$$\alpha > \alpha_* \triangleq \frac{1}{2\sqrt{1+s_*^2}} E\left(\frac{1}{1+s_*^2}\right) \approx 0.53, \tag{3.24}$$

where s_*^2 is the unique solution to

$$2E\left(\frac{1}{1+s_*^2}\right) = K\left(\frac{1}{1+s_*^2}\right).$$

Here, $K(\cdot)$ and $E(\cdot)$ denote the complete elliptic integrals introduced in Equation (3.11). Further, when $\alpha > \alpha_*$ and $\delta > \delta_{\text{AMP}}$, then $F_2(\sigma^2)$ is strongly globally attracting in $[0, \sigma^2_{\max}]$. Specifically,

$$\sigma^2 < \psi_2(\alpha, \sigma^2; \delta) < F_2(\alpha), \quad \sigma^2 \in (0, F_2(\alpha)),$$

and

$$F_2(\alpha) < \psi_2(\alpha, \sigma^2) < \sigma^2, \quad \sigma^2 \in (F_2(\alpha), \sigma^2_{\max}).$$

We prove this lemma in Appendix A.3.4.2.

Finally, we present the properties of F_1 and F_2 in the following lemma.

Lemma 3.11. The following hold for $F_1(\sigma^2)$ and $F_2(\alpha; \delta)$ (for $\delta > 2$):

(i) $F_1(0) = 1$ and $\lim_{\sigma^2 \to \frac{\pi^2}{16}-} F_1(\sigma^2) = 0$. Further, by choosing $F_1(\frac{\pi^2}{16}) = 0$, we have $F_1(\sigma^2)$ is continuous on $\left[0, \frac{\pi^2}{16}\right]$ and strictly decreasing in $\left(0, \frac{\pi^2}{16}\right)$;

(ii) F_2 is a continuous function of $\alpha \in [0, 1]$ and $\delta \in (2, \infty)$. $F_2(1; \delta) = 0$, and $F_2(0; \delta) = \left(\frac{-\pi + \sqrt{\pi^2 + 4(\delta-4)}}{\delta - 4}\right)^2$ for $\delta \neq 4$ and $F_2(0; 4) = 4/\pi^2$.

We prove this lemma in Appendix A.3.4.3.

With these properties we have proved for ψ_1, ψ_2, F_1 and F_2, we can prove Lemma 3.5 and details are presented in Appendix A.3.4.4.

3.3.3 Proof of Lemma 3.7

First, we introduce a function that will be crucial for our proof.

Definition 7. Define

$$L(\alpha;\delta) \triangleq \frac{4}{\delta}\left(1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4\left[1+(\phi_1^{-1}(\alpha))^2\right]}\right), \quad \alpha \in (0,1), \tag{3.25}$$

where $\phi_1 : \mathbb{R}_+ \mapsto [0,1]$ and $\phi_2 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ below:

$$\phi_1(s) \triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{3.26a}$$

$$\phi_2(s) \triangleq \int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta + s^2}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{3.26b}$$

where ϕ_1^{-1} is the inverse functions of ϕ_1. The existence of ϕ_1^{-1} follows from its monotonicity, which can be seen from its definition.

In the following, we list some preliminary properties of $L(\alpha;\delta)$. The main proof for Lemma 3.7 comes afterwards.

- **Preliminaries:**

  The following lemma helps us clarify the importance of L in the analysis of the dynamics of SE:

  Lemma 3.12. For any $\alpha > 0$, $\sigma^2 > 0$ and $\delta > 0$, the following holds:

$$L\left[\psi_1(\alpha,\sigma^2);\delta\right] \leq \psi_2(\alpha,\sigma^2;\delta), \tag{3.27}$$

where ψ_1 and ψ_2 are the SE maps defined in Equation (3.6), and $L(\alpha;\delta)$ is defined in Equation (3.25).

We prove this lemma in Appendix A.3.4.5

To understand the implication of this lemma, let us consider the t^{th} iteration of the SE:

$$\alpha_{t+1} = \psi_1(\alpha_t, \sigma_t^2),$$
$$\sigma_{t+1}^2 = \psi_2(\alpha_t, \sigma_t^2; \delta),$$

Note that according to Lemma 3.12, no matter where (α_t, σ_t^2) is, $(\alpha_{t+1}, \sigma_{t+1}^2)$ will fall above the $\sigma^2 = L(\alpha;\delta)$ curve. This function is a key component in the dynamics of AMP.A. More specifically, for models in Chapter 2, monotonicity of the update functions ensures that the estimates will move towards the fixed point in the corresponding direction but will not cross them. Therefore we can use rectangles to bound them. In the current case, because ψ_2 is not a monotonic function, the estimates (α_t, σ_t^2) can cross the fixed point on $F_2(\alpha)$ and can potentially oscillate between $\mathcal{R}_1$ and $\mathcal{R}_{2b}$ (instead of $\mathcal{R}_{2a}$) in Figure 3.6. The function L serves the purpose of preventing this kind of oscillation happens. Because of the importance of the function L, we discuss some main properties of the function $L(\alpha;\delta)$ before we proceed further.

Lemma 3.13. $L(\alpha;\delta)$ is a strictly decreasing function of $\alpha \in (0,1)$.

We prove this lemma in Appendix A.3.4.6

The next lemma compares the function $L(\alpha;\delta)$ with $F_1^{-1}(\alpha)$.

Lemma 3.14. If $\delta > \delta_{\text{AMP}}$, then

$$F_1^{-1}(\alpha) > L(\alpha;\delta), \quad \forall \alpha \in (0,1).$$

We prove this lemma in Appendix A.3.4.7

The third lemma shows the lower bound on $L(\alpha;\delta)$ for all $\alpha \in (0,1)$ and $\delta > 0$.

Lemma 3.15. The following holds for any $\alpha \in (0,1)$ and $\delta > 0$,

$$L(\alpha;\delta) > \frac{4}{\delta}\left(1 - \frac{\pi^2}{16} - \frac{1}{2}\alpha^2\right), \tag{3.28}$$

where $L(\alpha,\delta)$ is defined in Equation (3.25).

We prove this lemma in Appendix A.3.4.8.

Finally, we analyze the monotonic properties of ψ_2 for $\sigma^2 \in [L(\alpha;\delta_{\text{AMP}}),\infty)$ and at $\sigma^2 = L(\alpha;\delta)$ in the following two lemmas.

Lemma 3.16. For any $\alpha \in [0,1]$, $\psi_2(\alpha,\sigma^2;\delta_{\text{AMP}})$ is an increasing function of σ^2 on $\sigma^2 \in [L(\alpha;\delta_{\text{AMP}}),\infty)$, where the function $L(\alpha;\delta)$ is defined in Equation (7).

We prove this lemma in Appendix A.3.4.9.

Lemma 3.17. For any $\alpha \in [0,1]$, $\psi_2\left(\alpha, L(\alpha;\delta);\delta\right)$ is a strictly decreasing function of $\delta > 0$, where $L(\alpha;\delta)$ is defined in Equation (3.25).

We prove this lemma in Appendix A.3.4.10.

- **Main proof**

We now return to the main proof for Lemma 3.7. Notice that by Lemma 3.12, $(\alpha_{t_0},\sigma^2_{t_0})$ cannot fall below the curve $L(\alpha;\delta)$ for $t_0 \geq 1$. Hence, for $\mathcal{R}_2$, we can focus on the region above $L(\alpha;\delta)$ (including $L(\alpha;\delta)$), which we denote as $\mathcal{R}_{2a}$. See Fig. 3.6 for illustration.

We will first prove that if $(\alpha,\sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$, then the next iterates $\psi_1(\alpha,\sigma^2)$

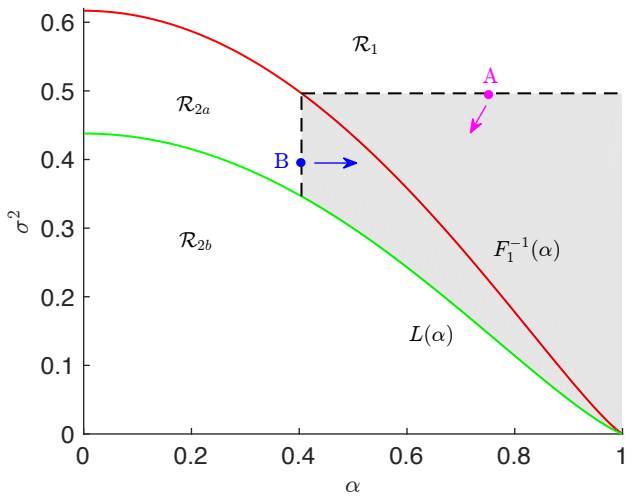

Figure 3.6: Illustration of the convergence behavior. $\mathcal{R}_1$ and $\mathcal{R}_2$ are defined in Definition 6. For both point A and point B, $B_1(\alpha, \sigma^2)$ and $B_2(\alpha, \sigma^2)$ are given by the two dashed lines. After one iteration, $\mathcal{R}_{2b}$ will not be achievable and we can focus on $\mathcal{R}_{2a}$.

and $\psi_2(\alpha, \sigma^2)$ satisfy the following:

$$\psi_1(\alpha, \sigma^2) \geq B_1(\alpha, \sigma^2), \tag{3.29a}$$

and

$$\psi_2(\alpha, \sigma^2) \leq B_2(\alpha, \sigma^2), \tag{3.29b}$$

where $B_1(\alpha, \sigma^2)$ and $B_2(\alpha, \sigma^2)$ are defined as

$$\begin{aligned} B_1(\alpha, \sigma^2) &\triangleq \min\left\{\alpha, F_1(\sigma^2)\right\}, \\ B_2(\alpha, \sigma^2) &\triangleq \max\left\{\sigma^2, F_1^{-1}(\alpha)\right\}. \end{aligned} \tag{3.30}$$

Note that when (α, σ^2) is on F_1^{-1} (i.e., $\sigma^2 = F_1^{-1}(\alpha)$), equalities in Equation (3.29a) and Equation (3.29b) can be achieved. Further, this is the only case when either of the equality is achieved. Also, it is easy to see that if (α, σ^2) is

on F_1^{-1}, then $(\psi_1(\alpha,\sigma^2),\psi_2(\alpha,\sigma^2))$ cannot be on F_1^{-1}.

Since F_1^{-1} separates $\mathcal{R}_1$ and $\mathcal{R}_{2a}$, Equation (3.30) can also be written as

$$\big[B_1(\alpha,\sigma^2),B_2(\alpha,\sigma^2)\big]=\begin{cases}[F_1(\sigma^2),\sigma^2] & \text{if } (\alpha,\sigma^2)\in\mathcal{R}_1,\\ [\alpha,F_1^{-1}(\alpha)] & \text{if } (\alpha,\sigma^2)\in\mathcal{R}_{2a}.\end{cases}\tag{3.31}$$

As a concrete example, consider the situation shown in Fig. 3.6. In this case, for both point A and point B, $B_1(\alpha,\sigma^2)$ and $B_2(\alpha,\sigma^2)$ are given by the two dashed lines. This directly follows from Equation (3.31) by noting that point A is in region $\mathcal{R}_1$ and point B is in region $\mathcal{R}_{2a}$. Let $\mathcal{R}_{2a}\backslash F_1^{-1}(\alpha)$ be a shorhand for $\{(\alpha,\sigma^2)|(\alpha,\sigma^2)\in\mathcal{R}_{2a},\alpha\neq F_1(\sigma^2)\}$. To prove the strict inequality in Equation (3.29), we deal with $(\alpha,\sigma^2)\in\mathcal{R}_1$ and $(\alpha,\sigma^2)\in\mathcal{R}_{2a}\backslash F_1^{-1}(\alpha)$ separately.

1. Assume that $(\alpha,\sigma^2)\in\mathcal{R}_1$. Using Equation (3.31), the inequality in Equation (3.29) can be rewritten as

$$\psi_1(\alpha,\sigma^2)>F_1(\sigma^2)\quad\text{and}\quad\psi_2(\alpha,\sigma^2)<\sigma^2.\tag{3.32}$$

   Since $(\alpha,\sigma^2)\in\mathcal{R}_1$, we have $\sigma^2>F_1^{-1}(\alpha)$. Then, applying Equation (3.21) proves $\psi_1(\alpha,\sigma^2)>F_1(\sigma^2)$. Further, using Lemma 3.5, we have $\sigma^2>F_1^{-1}(\alpha)>F_2(\alpha)$. Also, Lemma 3.6 guarantees that $\sigma^2<\sigma^2_{\max}$. Hence, $F_1^{-1}(\alpha)<\sigma^2<\sigma^2_{\max}$ and applying Lemma 3.10 (iv) yields $\psi_2(\alpha,\sigma^2)<\sigma^2$.

2. We now consider the case where $(\alpha,\sigma^2)\in\mathcal{R}_{2a}\backslash F_1^{-1}(\alpha)$. Similar to Equation (3.32), we need to prove

$$\psi_1(\alpha,\sigma^2)>\alpha\quad\text{and}\quad\psi_2(\alpha,\sigma^2)<F_1^{-1}(\alpha).\tag{3.33}$$

The inequality $\psi_1(\alpha, \sigma^2) > \alpha$ can be proved by the global attractiveness in Lemma 3.9 (iii) and the fact that $\sigma^2 < F_1^{-1}(\alpha)$ when $(\alpha, \sigma^2) \in \mathcal{R}_{2a} \backslash F_1^{-1}(\alpha)$. The proof for $\psi_2(\alpha, \sigma^2) < F_1^{-1}(\alpha)$ is considerably more complicated and is detailed in Lemma 3.18 below.

Lemma 3.18. For any $(\alpha, \sigma^2) \in \mathcal{R}_{2a}$ (see Definition 6) and $\delta \geq \delta_{\text{AMP}}$, the following holds:

$$\psi_2(\alpha, \sigma^2; \delta) < F_1^{-1}(\alpha), \tag{3.34}$$

where ψ_2 is the SE map in Equation (3.6b) and F_1^{-1} is the inverse of F_1 defined in Lemma 3.9.

Proof. The following holds when $(\alpha, \sigma^2) \in \mathcal{R}_{2a}$:

$$\psi_2(\alpha, \sigma^2; \delta) \leq \max_{\hat{\sigma}^2 \in \mathcal{D}_\alpha} \psi_2(\alpha, \hat{\sigma}^2; \delta),$$

where

$$\mathcal{D}_\alpha \triangleq \left\{ \hat{\sigma}^2 \middle| L(\alpha; \delta) \leq \sigma^2 \leq F_1^{-1}(\alpha) \right\}. \tag{3.35}$$

Hence, to prove Equation (3.34), it suffices to prove that the following holds for any $\delta \geq \delta_{\text{AMP}}$ and $\alpha \in [0, 1]$:

$$\max_{\hat{\sigma}^2 \in \mathcal{D}_\alpha} \psi_2(\alpha, \hat{\sigma}^2; \delta) < F_1^{-1}(\alpha). \tag{3.36}$$

We next prove Equation (3.36). We consider the three different cases:

(i) $\alpha \in [\alpha_*, 1]$ and all $\delta \in [\delta_{\text{AMP}}, \infty)$, where α_* is defined in Equation (3.24).

(ii) $\alpha \in [0, \alpha_*)$ and $\delta \in [\delta_{\text{AMP}}, 17]$.

(iii) $\alpha \in [0, \alpha_*)$ and $\delta \in (17, \infty)$.

Case (i): Lemma 3.10 (v) shows that ψ_2 is an increasing function of σ^2 in $\mathbb{R}_+$. Hence, by noting Equation (3.35), we have

$$\max_{\hat{\sigma}^2 \in \mathcal{D}_\alpha} \psi_2(\alpha, \hat{\sigma}^2; \delta) = \psi_2(\alpha, F_1^{-1}(\alpha); \delta).$$

Therefore, proving Equation (3.40) reduces to proving

$$\psi_2(\alpha, F_1^{-1}(\alpha); \delta) \leq F_1^{-1}(\alpha). \tag{3.37}$$

Finally, Equation (3.37) follows from the global attractiveness property in Lemma 3.10 (iv) and the inequality $F_1^{-1}(\alpha) > F_2(\alpha; \delta)$ in Lemma 3.5.

Case (ii): We will prove that the following holds for $\alpha \in [0, \alpha_*)$ and $\delta \in [\delta_{\text{AMP}}, 17]$ (at the end of this proof)

$$\max_{\hat{\sigma}^2 \in \mathcal{D}_\alpha} \psi_2(\alpha, \sigma^2; \delta) = \max\left\{\psi_2(\alpha, L(\alpha; \delta); \delta),\ \psi_2(\alpha, F_1^{-1}(\alpha); \delta)\right\}. \tag{3.38}$$

Namely, the maximum of ψ_2 over σ^2 is achieved at either $\sigma^2 = L(\alpha; \delta)$ or $\sigma^2 = F_1^{-1}(\alpha)$. Hence, we only need to prove that the following holds for any $\alpha \in [0, \alpha_*)$ and $\delta \geq \delta_{\text{AMP}}$:

$$\max\left\{\psi_2(\alpha, L(\alpha; \delta); \delta),\ \psi_2(\alpha, F_1^{-1}(\alpha); \delta)\right\} \leq F_1^{-1}(\alpha). \tag{3.39}$$

In the sequel, we first use Equation (3.38) to prove Equation (3.36), and the proof for Equation (3.38) will come at the end of this proof.

Firstly, it is easy to see that $\psi_2(\alpha, F_1^{-1}(\alpha); \delta)$ is a decreasing function of δ, since $\psi_2(\alpha, \sigma^2; \delta)$ is a decreasing function of δ and $F_1^{-1}(\alpha)$ does not depend on δ. Further, Lemma 3.17 shows that $\psi_2(\alpha, L(\alpha; \delta); \delta)$ is also a decreasing

function of δ. (Notice that unlike $F_1^{-1}(\alpha)$, $L(\alpha;\delta)$ depends on δ, and thus Lemma 3.17 is nontrivial.) Hence, to prove Equation (3.39) for $\delta \geq \delta_{\mathrm{AMP}}$, it suffices to prove Equation (3.39) for $\delta = \delta_{\mathrm{AMP}}$, namely,

$$\max\left\{\psi_2(\alpha, L(\alpha;\delta);\delta_{\mathrm{AMP}}),\ \psi_2(\alpha, F_1^{-1}(\alpha);\delta_{\mathrm{AMP}})\right\} \leq F_1^{-1}(\alpha). \tag{3.40}$$

When $\delta = \delta_{\mathrm{AMP}}$, we prove in Lemma 3.16 that ψ_2 is an increasing function of σ^2 in $\sigma^2 \in [L(\alpha;\delta_{\mathrm{AMP}}), \infty)$. (Such monotonicity generally does not hold if δ is too large.) Further, Lemma 3.14 shows that $F_1^{-1}(\alpha) > L(\alpha;\delta_{\mathrm{AMP}})$. Hence,

$$\psi_2(\alpha, L(\alpha;\delta);\delta_{\mathrm{AMP}}) \leq \psi_2(\alpha, F_1^{-1}(\alpha);\delta_{\mathrm{AMP}}),$$

and thus proving Equation (3.40) reduces to proving

$$\psi_2(\alpha, F_1^{-1}(\alpha);\delta_{\mathrm{AMP}}) \leq F_1^{-1}(\alpha),$$

which follows from the same argument as that for Equation (3.37).

Case (iii): Lemma 3.10 (iii) shows that $\psi_2(\alpha;\sigma^2;\delta) \leq \frac{4}{\delta}$ for any $\sigma^2 \in [0, \sigma_{\max}^2]$. It is easy to see that $\mathcal{D}_\alpha \subset [0, \sigma_{\max}^2]$, and thus

$$\max_{\sigma^2 \in \mathcal{D}_\alpha} \psi_2(\alpha, \sigma^2;\delta) \leq \frac{4}{\delta} \leq \frac{4}{17} \approx 0.235. \tag{3.41}$$

Further, Lemma 3.11 shows that $F_1^{-1} : [0,1] \mapsto [0, \pi^2/16]$ is monotonically decreasing. Hence,

$$F_1^{-1}(\alpha) > F_1^{-1}(\alpha_*) \approx 0.415, \tag{3.42}$$

where the numerical constant is calculated from the closed form formula $F_1^{-1}(\alpha) = \alpha^2 \cdot \left[\phi_1^{-1}(\alpha)\right]^2$ (see Equation (A.135)) and $\alpha_* \approx 0.5274$ (from Equation (3.24)). Comparing Equation (3.41) and Equation (3.42) shows that Equation (3.36) holds in this case.

It only remains to prove Equation (3.38). We have shown in Equation (A.119) that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta \alpha}\left(\alpha - \underbrace{\frac{1}{2\sqrt{1+s^2}} E\left(\frac{1}{1+s^2}\right)}_{f(s)}\right), \tag{3.43}$$

where $s \triangleq \sigma/\alpha$. Further, we have proved in Equation (A.121) that $f(s)$ is strictly increasing on $[0, s_*)$ and strictly decreasing on (s_*, ∞), where s_* is defined in Equation (3.25). Hence, when $f(0) = 0.5 < \alpha < f(s_*) = \alpha_*$, there exist two solutions to

$$\alpha = f(s),$$

denoted as $s_1(\alpha)$ and $s_2(\alpha)$, respectively. Also, from Equation (3.43) and noting the definition $s = \sigma/\alpha$, we have

$$\begin{aligned}
\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} > 0 &\Longleftrightarrow \sigma^2 \in \left[0, \sigma_1^2(\alpha)\right) \cup \left(\sigma_2^2(\alpha), \infty\right), \\
\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} \leq 0 &\Longleftrightarrow \sigma^2 \in \left[\sigma_1^2(\alpha), \sigma_2^2(\alpha)\right],
\end{aligned}$$

where $\sigma_1^2(\alpha) \triangleq \alpha^2 s_1^2(\alpha)$ and $\sigma_2^2(\alpha) \triangleq \alpha^2 s_2^2(\alpha)$. Hence, for fixed α where $\alpha \in (f(0), f(s_*))$, $\sigma_1^2(\alpha)$ is a local maximum of ψ_2 and $\sigma_2^2(\alpha)$ is a local minimum. Clearly, if

$$L(\alpha; \delta) \geq \sigma_1^2(\alpha), \tag{3.44}$$

then the maximum of ψ_2 over $\sigma^2 \in [L(\alpha;\delta), F_1^{-1}(\alpha)]$ can only happen at either $L(\alpha;\delta)$ or $F_1^{-1}(\alpha)$, which will prove Equation (3.38). Further, for the degenerate case $\alpha \in (0, f(0))$, ψ_2 only has a local minimum, and it is easy to see that Equation (3.38) also holds. Thus, we only need to prove that Equation (3.44) holds when $\delta < 17$. This can be proved as follows:

$$\sigma_1^2(\alpha) \overset{(a)}{\leq} s_*^2 \cdot \alpha^2 \overset{(b)}{\leq} s_*^2 \cdot \alpha_*^2, \tag{3.45}$$

where (a) is from the fact that $s_1(\alpha) \leq s_*$ and (b) is from our assumption $\alpha \leq \alpha_*$. On the other hand, since $L(\alpha)$ is a decreasing function of α (see Lemma 3.13), and thus for $\alpha \leq \alpha_*$ we have

$$\begin{aligned} L(\alpha;\delta) &\geq L(\alpha_*;\delta) \\ &= \frac{4}{\delta}\left(1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha_*))}{4\left[1+(\phi_1^{-1}(\alpha_*))^2\right]}\right), \end{aligned} \tag{3.46}$$

where the last step is from Definition 3.25. Based on Equation (3.45) and Equation (3.46), we see that $L(\alpha;\delta) > \sigma_1^2(\alpha)$ for $\alpha \leq \alpha_*$ if

$$\delta \leq \frac{4}{s_*^2 \cdot \alpha_*^2}\left(1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha_*))}{4\left[1+(\phi_1^{-1}(\alpha_*))^2\right]}\right) \approx 17.04,$$

where the numerical constant is calculated based on the definition of α_* in Equation (A.125), the definition of s_* in Equation (3.25), and that of ϕ_1 and ϕ_2 in Definition 3.25. Hence, the condition $\delta < 17$ is enough for our purpose. This concludes our proof. □

Now we turn our attention to the proof of part (i) of Lemma 3.7. Suppose that $(\alpha, \sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$. Then, using Equation (3.29) and based on the fact that

$F_1(\alpha)$ is a strictly decreasing function, we know that $(\psi_1(\alpha,\sigma^2),\psi_2(\alpha,\sigma^2)) \in \mathcal{R}_1 \cup \mathcal{R}_2$ (See Definition 6). Further, Lemma 3.8 shows that $(\psi_1(\alpha,\sigma^2),\psi_2(\alpha,\sigma^2)) \notin \mathcal{R}_{2b}$. Hence, $(\psi_1(\alpha,\sigma^2),\psi_2(\alpha,\sigma^2)) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$. Applying this argument recursively shows that if $(\alpha_{t_0},\sigma_{t_0}^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$, then $(\alpha_t,\sigma_t^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$ for all $t > t_0$. An illustration of the situation is shown in Fig. 3.6.

Now we can discuss the proof of part (ii) of Lemma 3.7. To proceed, we introduce two auxiliary sequences $\{\tilde{\alpha}_{t+1}\}_{t\geq t_0}$ and $\{\tilde{\sigma}_{t+1}^2\}_{t\geq t_0}$, defined as:

$$\tilde{\alpha}_{t+1} = B_1(\alpha_t,\sigma_t^2) \quad \text{and} \quad \tilde{\sigma}_{t+1}^2 = B_2(\alpha_t,\sigma_t^2), \tag{3.47}$$

where B_1 and B_2 are defined in Equation (3.30). Note that the definitions of $B_1(\alpha,\sigma^2)$ and $B_2(\alpha,\sigma^2)$ require $(\alpha,\sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$, and such requirement is satisfied here due to part (i) of this lemma. Noting the SE update $\alpha_{t+1} = \psi_1(\alpha_t,\sigma_t^2)$ and $\sigma_{t+1}^2 = \psi_2(\alpha_t,\sigma_t^2)$, and recall the inequalities in Equation (3.29), we obtain the following:

$$\alpha_{t+1} \geq \tilde{\alpha}_{t+1} \quad \text{and} \quad \sigma_{t+1}^2 \leq \tilde{\sigma}_{t+1}^2, \quad \forall t \geq t_0. \tag{3.48}$$

Namely, $\{\tilde{\alpha}_{t+1}\}_{t\geq t_0}$ and $\{\tilde{\sigma}_{t+1}^2\}_{t\geq t_0}$ are "worse" than $\{\alpha_{t+1}\}_{t\geq t_0}$ and $\{\sigma_{t+1}^2\}_{t\geq t_0}$, respectively, at each iteration. We next prove that

$$\lim_{t\to\infty} \tilde{\alpha}_{t+1} = 1 \quad \text{and} \quad \lim_{t\to\infty} \tilde{\sigma}_{t+1}^2 = 0, \tag{3.49}$$

which together with Equation (3.48), and the fact that $\alpha_{t+1} \leq 1$ and $\sigma_{t+1} > 0$

(since $(\alpha_t, \sigma_t^2) \in \mathcal{R}_{2a}$), leads to the results we want to prove:

$$\lim_{t\to\infty} \alpha_{t+1} = 1 \quad \text{and} \quad \lim_{t\to\infty} \sigma_{t+1}^2 = 0.$$

It remains to prove Equation (3.49). First, notice that $\tilde{\alpha}_{t+1} \leq 1$ and $\tilde{\sigma}_{t+1}^2 \geq 0$ ($\forall t \geq t_0$), from the definition in Equation (3.30). We then show that the sequence $\{\tilde{\alpha}_{t+1}\}_{t\geq t_0}$ is monotonically non-decreasing and $\{\tilde{\sigma}_{t+1}^2\}_{t\geq t_0}$ is monotonically non-increasing, namely,

$$\tilde{\alpha}_{t+2} \geq \tilde{\alpha}_{t+1} \quad \text{and} \quad \tilde{\sigma}_{t+2}^2 \leq \tilde{\sigma}_{t+1}^2, \quad \forall t \geq t_0, \tag{3.50}$$

and equalities of Equation (3.50) hold only when the equalities in Equation (3.29) hold. Then we can finish the proof by the fact that $\tilde{\alpha}$ and $\tilde{\sigma}^2$ will improve strictly in at most two consecutive iterations and the ratios $\frac{\tilde{\alpha}_{t+2}}{\tilde{\alpha}_t}, \frac{\tilde{\sigma}_{t+2}^2}{\tilde{\sigma}_t^2}$ are continuous functions of (α_t, σ_t^2) on $[\tilde{\alpha}_{t_0}, 1] \times [0, \sigma_{\max}^2]$. (This is essentially due to the fact that equalities in Equation (3.29) can be achieved when $\sigma^2 = F_1^{-1}(\alpha)$, but this cannot happen in two consecutive iterations. See the discussions below Equation (3.30).)

To prove Equation (3.50), we only need to prove the following (based on the definition in Equation (3.47))

$$B_1[\psi_1, \psi_2] \geq B_1(\alpha, \sigma^2) \quad \text{and} \quad B_2[\psi_1, \psi_2] \leq B_2(\alpha, \sigma^2), \quad \forall (\alpha, \sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a},$$

where ψ_1 and ψ_2 are shorthands for $\psi_1(\alpha, \sigma^2)$ and $\psi_2(\alpha, \sigma^2; \delta)$. From Equation

(3.30), the above inequalities are equivalent to

$$\min\{\psi_1, F_1(\psi_2)\} \geq B_1(\alpha, \sigma^2), \tag{3.51}$$

and

$$\max\left\{\psi_2, F_1^{-1}(\psi_1)\right\} \leq B_2(\alpha, \sigma^2). \tag{3.52}$$

Note that Equation (3.29) already proves the following

$$\psi_1 \geq B_1(\alpha, \sigma^2) \quad \text{and} \quad \psi_2 \leq B_2(\alpha, \sigma^2).$$

Hence, to prove Equation (3.51) and Equation (3.52), we only need to prove

$$F_1(\psi_2) \geq B_1(\alpha, \sigma^2) \quad \text{and} \quad F_1^{-1}(\psi_1) \leq B_2(\alpha, \sigma^2).$$

To prove $F_1(\psi_2) \geq B_1(\alpha, \sigma^2)$, we note that

$$\begin{aligned}
\psi_2 &\overset{(a)}{\leq} B_2(\alpha, \sigma^2) \\
&\overset{(b)}{=} \max\left\{\sigma^2, F_1^{-1}(\alpha)\right\} \\
&\overset{(c)}{=} F_1^{-1}\left(\min\left\{F_1(\sigma^2), \alpha\right\}\right) \\
&\overset{(d)}{=} F_1^{-1}\left(B_1(\alpha, \sigma^2)\right),
\end{aligned}$$

where (a) is from Equation (3.29b), (b) is from Equation (3.30), and (c) is due to the fact that F_1^{-1} is strictly decreasing, and (d) from Equation (3.29). Hence, since F_1 is strictly decreasing, we have

$$F_1(\psi_2) \geq F_1\left[F_1^{-1}\left(B_1(\alpha, \sigma^2)\right)\right] = B_1(\alpha, \sigma^2).$$

Further, it is straightforward to see that if both inequalities are strict in Equation (3.29) then

$$\min\{\psi_1, F_1(\psi_2)\} > B_1(\alpha, \sigma^2).$$

This shows that equalities of Equation (3.50) hold only when the equalities in Equation (3.29) hold.

The proof for $F_1^{-1}(\psi_1) \leq B_2(\alpha, \sigma^2)$ is similar and omitted.

3.3.4 Proof of Lemma 3.8

Suppose that $(\alpha, \sigma^2) \in \mathcal{R}_0$. From Definition 6, we have

$$\frac{\pi^2}{16} < \sigma^2 \leq \sigma_{\max}^2. \tag{3.53}$$

Further, F_1^{-1} is monotonically decreasing and hence (for $\delta > \delta_{\text{AMP}}$)

$$\frac{\pi^2}{16} = F_1^{-1}(0) > F_1^{-1}(\alpha) \geq F_2(\alpha; \delta), \tag{3.54}$$

where the last inequality is due to Lemma 3.5. Combining Equation (3.53) and Equation (3.54) yields

$$F_2(\alpha; \delta) < \sigma^2 \leq \sigma_{\max}^2. \tag{3.55}$$

By the global attractiveness property in Lemma 3.10 (iv), Equation (3.55) implies

$$\psi_2(\alpha; \sigma^2; \delta) < \sigma^2.$$

From the above analysis, we see that as long as $\frac{\pi^2}{16} < \sigma_t^2 \leq \sigma_{\max}^2$ (and also $0 <$

$\alpha_t < 1$), σ_{t+1}^2 will be strictly smaller than σ_t^2:

$$\sigma_{t+1}^2 = \psi_2(\alpha_t; \sigma_t^2; \delta) < \sigma_t^2.$$

Hence, there exists a finite number $T \geq 1$ such that

$$\sigma_{T-1}^2 > \frac{\pi^2}{16} \quad \text{and} \quad \sigma_T^2 \leq \frac{\pi^2}{16}.$$

Otherwise, σ_t^2 will converge to a $\bar{\sigma}^2$ in $\mathcal{R}_0$. This implies that $\bar{\sigma}^2$ is a fixed point of ψ_2 for certain value of $0 < \alpha \leq 1$. However, we know from part (i) of Lemma 3.11 and Lemma 3.5 that this cannot happen.

Based on a similar argument, we also have $\psi_1(\alpha; \sigma^2) < \alpha$ and so $\alpha_{t+1} < \alpha_t$ for $t \leq T-1$. Further, we can show that $\alpha_t > 0$ (i.e., $\alpha_t \neq 0$) for all $0 \leq t \leq T$. First, $\alpha_0 > 0$ follows from our assumption. Further, from Equation (3.6a) we see that $\alpha_{t+1} > 0$ if $\alpha_t > 0$. Then, using a simple induction argument we prove that $\alpha_t > 0$ for all $0 \leq t \leq T$. Putting things together, we showed that there exists a finite number $T \geq 1$ such that

$$0 < \alpha_T \leq 1 \quad \text{and} \quad \sigma_T^2 \leq \frac{\pi^2}{16}.$$

(Recall that we have proved in Lemma 3.6 that $\alpha_T \leq 1$.) From Definition 6, $(\alpha_T, \sigma_T^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$.

3.4 Proof of Theorem 3.3

We consider the two different cases separately: (1) $\delta > \delta_{\text{global}}$ and (2) $\delta < \delta_{\text{global}}$.

3.4.1 Case $\delta > \delta_{\text{global}}$

In this section, we will prove that when $\delta > \delta_{\text{global}}$ the state evolution converges to the fixed point $(\alpha, \sigma^2) = (1, 0)$ if initialized close enough to the fixed point. We first claim the following lemma, which shows that F_1^{-1} is larger than $F_2(\alpha; \delta)$ for α close to one.

Lemma 3.19. Suppose that $\delta > \delta_{\text{global}} = 2$. Then, there exists an $\epsilon > 0$ such that the following holds:

$$F_1^{-1}(\alpha) > F_2(\alpha; \delta), \quad \forall \alpha \in (1 - \epsilon, 1). \tag{3.56}$$

We prove this lemma in Appendix A.3.5.

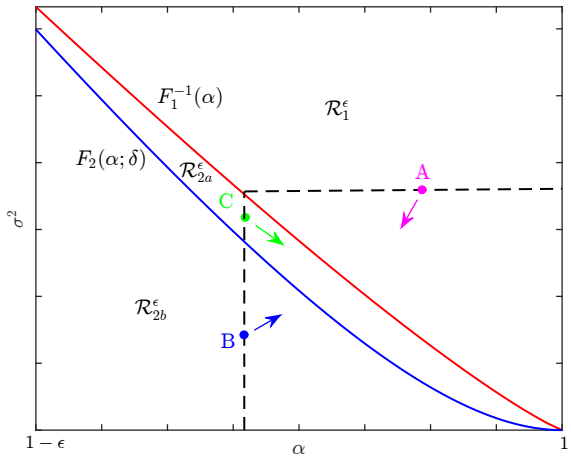

Figure 3.7: Illustration of the local convergence behavior when $\delta > \delta_{\text{global}}$. For all the three points shown in the figure, B_1 and B_2 are given by the dashed lines.

The idea of the proof is similar to that of Theorem 3.2. There are some differences though, since now δ can be smaller than δ_{AMP} and some results in the proof of Theorem 3.2 do not hold for the case considered here. On the other hand, as we focus on the range $\alpha \in (1-\epsilon, 1) > \alpha_*$, and under this condition we know that $F_2(\sigma^2; \delta)$ is strongly globally attracting (see Lemma 3.10-(v)), which means that $\psi_2(\alpha, \sigma^2)$ moves towards the fixed point $F_2(\alpha; \delta)$, but cannot move to the other side of $F_2(\alpha; \delta)$.

We continue to prove the local convergence of the state evolution. We divide the region $\mathcal{R}^\epsilon \triangleq \{(\alpha,\sigma^2)|1-\epsilon \le \alpha \le 1, 0 \le \sigma^2 \le F_1^{-1}(1-\epsilon)\}$ into the following sub-regions:

$$\begin{aligned}
\mathcal{R}_1^\epsilon &\triangleq \left\{(\alpha,\sigma^2)\middle|1-\epsilon \le \alpha \le 1, F_1^{-1}(\alpha) < \sigma^2 \le F_1^{-1}(1-\epsilon)\right\}, \\
\mathcal{R}_{2a}^\epsilon &\triangleq \left\{(\alpha,\sigma^2)\middle|1-\epsilon \le \alpha \le 1, F_2(\alpha;\delta) < \sigma^2 \le F_1^{-1}(\alpha)\right\} \\
\mathcal{R}_{2b}^\epsilon &\triangleq \left\{(\alpha,\sigma^2)\middle|1-\epsilon \le \alpha \le 1, 0 \le \sigma^2 \le F_2(\alpha;\delta)\right\}.
\end{aligned} \tag{3.57}$$

Similar to the proof of Lemma 3.7 discussed in Section 3.3.3, we will show that if $(\alpha,\sigma^2) \in \mathcal{R}^\epsilon$ then the new states (ψ_1,ψ_2) can be bounded as follows:

$$\psi_1(\alpha,\sigma^2) \ge B_1(\alpha,\sigma^2) \quad \text{and} \quad \psi_2(\alpha,\sigma^2) \le B_2(\alpha,\sigma^2), \quad \forall(\alpha,\sigma^2) \in \mathcal{R}^\epsilon, \tag{3.58}$$

where

$$B_1(\alpha,\sigma^2) = \min\left\{\alpha, F_1(\sigma^2)\right\} \quad \text{and} \quad B_2(\alpha,\sigma^2) = \max\left\{\sigma^2, F_1^{-1}(\alpha)\right\}.$$

Based on the strong global attractiveness of ψ_1 (Lemma 3.9-iii) and ψ_2 (Lemma 3.10-v) and the additional result Equation (3.22), it is straightforward to show the following:

$$\begin{aligned}
\psi_1(\alpha,\sigma^2) \ge F_1(\sigma^2) \quad &\text{and} \quad \psi_2(\alpha,\sigma^2) \le \sigma^2, \quad \forall(\alpha,\sigma^2) \in \mathcal{R}_1^\epsilon, \\
\psi_1(\alpha,\sigma^2) \ge \alpha \quad &\text{and} \quad \psi_2(\alpha,\sigma^2) \le \sigma^2, \quad \forall(\alpha,\sigma^2) \in \mathcal{R}_{2a}^\epsilon, \\
\psi_1(\alpha,\sigma^2) \ge \alpha \quad &\text{and} \quad \psi_2(\alpha,\sigma^2) \le F_2(\alpha;\delta), \quad \forall(\alpha,\sigma^2) \in \mathcal{R}_{2b}^\epsilon,
\end{aligned}$$

which, together with the definitions given in Equation (3.57) and the fact that $F_2(\alpha;\delta) < F_1^{-1}(\alpha)$ (cf. Lemma 3.19), proves Equation (3.58). The rest of the proof

follows that in Section 3.3.3. Namely, we construct two auxiliary sequences $\{\tilde{\alpha}_{t+1}\}$ and $\{\tilde{\sigma}^2_{t+1}\}$ where

$$\tilde{\alpha}_{t+1} = B_1(\alpha_t, \sigma_t^2) \quad \text{and} \quad \tilde{\sigma}^2_{t+1} = B_2(\alpha_t, \sigma_t^2),$$

and show that $\{\tilde{\alpha}_{t+1}\}$ and $\{\tilde{\sigma}^2_{t+1}\}$ monotonically converge to 1 and 0 respectively. The detailed arguments can be found in Section 3.3.3 and will not be repeated here.

3.4.2 Case $\delta < \delta_{\text{global}}$

We proved in Equation (A.119) that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta \alpha}\left(\alpha - \underbrace{\frac{1}{2\sqrt{1+s^2}} E\left(\frac{1}{1+s^2}\right)}_{f(s)}\right),$$

where $s = \frac{\sigma}{\alpha}$. Hence, we have (note that $E(1) = 1$)

$$\partial_2 \psi_2(\alpha, 0) \triangleq \frac{\partial \psi_2(\alpha, \sigma^2)}{\partial \sigma^2}\Big|_{\sigma^2=0} = \frac{4}{\delta}\left(1 - \frac{1}{2\alpha}\right), \quad \forall \alpha > 0. \tag{3.59}$$

Therefore,

$$\partial_2 \psi_2(\alpha, 0) > 1, \quad \forall \alpha > \frac{2}{4-\delta}.$$

When $\delta < \delta_{\text{global}} = 2$, we have $\frac{2}{4-\delta} < 1$ and therefore there exists a constant α^* that satisfies the following:

$$\frac{2}{4-\delta} < \alpha^* < 1,$$

which together with Equation (3.59) yields

$$\partial_2 \psi_2(\alpha^*, 0) > 1.$$

Further, as discussed in the proof of Lemma 3.10-(i), $\partial_2\psi_2(\alpha^*, \sigma^2)$ is a continuous function of σ^2. Hence, there exists $\xi^* > 0$ such that

$$\partial_2\psi_2(\alpha^*, \sigma^2) > 1, \quad \forall \sigma^2 \in [0, \xi^*]. \tag{3.60}$$

Further, we have shown in Equation (A.112) that

$$\frac{\partial\psi_2(\alpha, \sigma^2; \delta)}{\partial\sigma^2} = \frac{4}{\delta}\left(1 - \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta\right),$$

and it is easy to see that $\partial_2\psi_2(\alpha, \sigma^2; \delta)$ is an increasing function of $\alpha \in (0, \infty)$. Hence, together with Equation (3.60) we get the following

$$\partial_2\psi_2(\alpha, \sigma^2; \delta) > 1, \quad \forall (\alpha, \sigma^2) \in [\alpha^*, 1] \times [0, \xi^*],$$

which means that $\psi_2(\alpha, \sigma^2) - \sigma^2$ is a strictly increasing function of σ^2 for $(\alpha, \sigma^2) \in [\alpha^*, 1] \times [0, \xi^*]$. Hence,

$$\psi_2(\alpha, \sigma^2) - \sigma^2 > \psi_2(\alpha, 0) = \frac{4}{\delta}(1 - \alpha)^2 \geq 0, \quad \forall (\alpha, \sigma^2) \in [\alpha^*, 1] \times [0, \xi^*].$$

This implies that σ^2 moves away from 0 in a neighborhood of the fixed point $(1, 0)$.

3.5 Proofs of Theorems 3.4

In light of Lemma 3.1, we assume that $\alpha_0 \geq 0$ throughout this section.

3.5.1 Discussion

The goal of this section is to prove Theorems 3.4. The strategy is similar to the proof of Theorem 3.2. We first construct the functions F_1^{-1} and F_2. Then, we show that these two functions will intersect at exactly one point when $\delta > \delta_{\text{AMP}}$. Finally, we discuss the dynamics of the state evolution and show that (α_t, σ_t^2) converge to the intersection of F_1^{-1} and F_2. However, there are a few differences that make the proof of the noisy case more challenging:

1. Recall that in the noiseless case, the curve F_1^{-1} is entirely above F_2 (except for the fixed point $(1, 0)$) if $\delta > \delta_{\text{AMP}}$. See the plot in Fig. 3.5. On the other hand, when there is some noise, the curve F_2 will move up a little bit (while F_1^{-1} is unchanged) and will cross F_1 at a certain $\alpha_\star \in (0, 1)$. As shown in Fig. A.3, F_1^{-1} is above F_2 for $\alpha < \alpha_\star$ and is below F_2 when $\alpha > \alpha_\star$.

2. In the noisy setting the dynamic of SE becomes more challenging. In fact (α_t, σ_t^2) can move in any direction around the fixed point. That makes the proof of convergence of (α_t, σ_t^2) more complicated.

3. In the noiseless setting the location of the fixed point of SE was $(\alpha, \sigma^2) = (1, 0)$. This is not the case for the noisy settings where the location of the fixed point depends on the noise variance.

In the sections below we go over the entire proof, but will skip the parts that are similar to the proof of the noiseless setting which was discussed in Section 3.3.

3.5.2 Preliminaries

In the noisy setting, $\psi_1(\alpha;\sigma^2)$ remains unchanged, and $\psi_2(\alpha,\sigma^2;\delta)$ is replaced by $\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2)$ below:

$$\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2) = \psi_2(\alpha,\sigma^2;\delta) + 4\sigma_w^2 \tag{3.61a}$$

$$= \frac{4}{\delta}\left\{\alpha^2+\sigma^2+1-\alpha\left[\phi_1\left(\frac{\sigma}{\alpha}\right)+\phi_3\left(\frac{\sigma}{\alpha}\right)\right]\right\}+4\sigma_w^2, \tag{3.61b}$$

where

$$\begin{aligned}\phi_1(s) &\triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta+s^2\right)^{\frac{1}{2}}}\mathrm{d}\theta,\\ \phi_3(s) &\triangleq \int_0^{\frac{\pi}{2}} \left(\sin^2\theta+s^2\right)^{\frac{1}{2}}\mathrm{d}\theta.\end{aligned} \tag{3.62}$$

Before we proceed to the analysis of ψ_1,ψ_2,F_1, and F_2, we list a few identities for ϕ_1 and ϕ_3 which will be used in our proofs later.

Lemma 3.20. ϕ_1 and ϕ_3 satisfy the following properties:

$$\begin{aligned}\phi_1(s) &= \frac{(1+s^2)E\left(\frac{1}{1+s^2}\right)-s^2K\left(\frac{1}{1+s^2}\right)}{\sqrt{1+s^2}},\\ \phi_3(s) &= \sqrt{1+s^2}E\left(\frac{1}{1+s^2}\right),\\ \phi_1(0) &= 1,\\ \left.\frac{\mathrm{d}\phi_1(s)}{\mathrm{d}s^2}s^2\right|_{s=0} &= \left.\frac{s^2(E-K)}{2\sqrt{1+s^2}}\right|_{s=0}=0,\\ \left.\frac{\mathrm{d}\phi_1(s)\phi_3(s)}{\mathrm{d}s^2}\right|_{s=0} &= \frac{1}{2}\left.\left(\frac{(1+s^2)E^2-s^2K^2}{1+s^2}\right)^2\right|_{s=0}=\frac{1}{2},\end{aligned} \tag{3.63}$$

where E and K are shorthands for $E\left(\frac{1}{1+s^2}\right)$ and $K\left(\frac{1}{1+s^2}\right)$ respectively in the last two identities.

The proof of this lemma is a simple application of the identities we derived in Section 3.1.4, and is hence skipped.

Our next lemma summarizes the main properties of ψ_1, ψ_2, F_1 and F_2 in the noisy phase retrieval problem.

Lemma 3.21. Let $\tilde{\sigma}^2_{\max} \triangleq \sigma^2_{\max} + 4\sigma_w^2$, where $\sigma^2_{\max} = \max\{1, 4/\delta\}$. For any $\delta > \delta_{\text{AMP}}$, there exists $\epsilon > 0$ such that when $0 < \sigma_w^2 < \epsilon$ the following statements hold simultaneously:

(a) For $0 \leq \alpha \leq 1$, we have $\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) \leq \tilde{\sigma}^2_{\max}$, $\forall \sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$.

(b) For $0 \leq \alpha \leq 1$, $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta) + 4\sigma_w^2$ admits a unique globally attracting fixed point, denoted as $F_2(\alpha; \delta, \sigma_w^2)$, in $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$. Further, if $\alpha \geq \alpha_*$ (note that $\alpha_* \approx 0.53$ is defined in Equation (3.24)), then $F_2(\alpha; \delta, \sigma_w^2)$ is strongly globally attractive. Finally, $F_2(\alpha; \delta, \sigma_w^2)$ is a continuous function of σ_w^2.

(c) The equation $F_1^{-1}(\alpha) = F_2(\alpha; \delta, \sigma_w^2)$ has a unique nonzero solution in $\alpha \in [0, 1]$. Let $\alpha_\star(\delta, \sigma_w^2)$ be that unique solution. Then, $F_1^{-1}(\alpha) > F_2(\alpha; \delta, \sigma_w^2)$ for $0 \leq \alpha < \alpha_\star(\delta, \sigma_w^2)$ and $F_1^{-1}(\alpha) < F_2(\alpha; \delta, \sigma_w^2)$ for $\alpha_\star(\delta, \sigma_w^2) < \alpha \leq 1$.

(d) There exists $\hat{\alpha}(\delta, \sigma_w^2)$, such that $F_2(\alpha; \delta, \sigma_w^2)$ is a strictly decreasing function on $\alpha \in (0, \hat{\alpha}(\delta, \sigma_w^2))$ and a strictly increasing function on $(\hat{\alpha}(\delta, \sigma_w^2), 1)$. Further, $\alpha_\star(\delta, \sigma_w^2) < \hat{\alpha}(\delta, \sigma_w^2) < 1$.

(e) Define $L(\alpha; \delta, \sigma_w^2) \triangleq L(\alpha; \delta) + 4\sigma_w^2$, where $L(\alpha; \delta)$ is defined in Equation (3.25). Then, $L(\alpha; \delta, \sigma_w^2) < F_1^{-1}(\alpha)$ for all $\alpha \in (0, \alpha_*]$, where $\alpha_* \approx 0.53$ is defined in Equation (3.24).

(f) For any $\alpha \in (0, \alpha_*]$ and $\sigma^2 \in [L(\alpha; \delta, \sigma_w^2), F_1^{-1}(\alpha)]$, we have $\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) \triangleq \psi_2(\alpha, \sigma^2; \delta) + 4\sigma_w^2 < F_1^{-1}(\alpha)$.

(g) $F_2(1;\delta,\sigma_w^2) < F_1^{-1}(\alpha_*)$.

We prove this lemma in Appendix A.3.7

3.5.3 Convergence of the SE

Our next lemma proves that the state evolution still converges to the desired fixed point for $0 < \alpha_0 \leq 1$ and $\sigma_0^2 \leq 1$ if $\delta > \delta_{\text{AMP}}$.

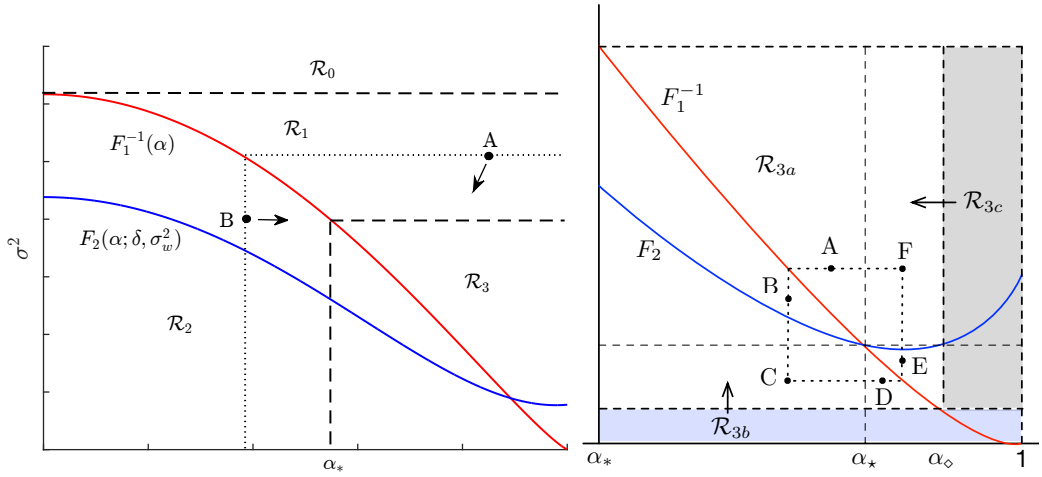

Figure 3.8: Dynamical behavior the state evolution in the low noise regime. **Left:** points in $\mathcal{R}_1$ and $\mathcal{R}_2$ will eventually move to $\mathcal{R}_3$. Here, $\alpha_* \approx 0.53$. **Right:** Illustration of $\mathcal{R}_3$. Points in $\mathcal{R}_{3b}$ and $\mathcal{R}_{3c}$ will eventually move to $\mathcal{R}_{3a}$. For points in $\mathcal{R}_{3a}$ (marked A, B, C, D, E, F), we can form a small rectangular region that bounds the remaining trajectory. Note that the lower and right bounds for A and B (and also the upper and left bounds for D and E) are given by $\sigma_\star^2$ and $\alpha_\star$ respectively.

Lemma 3.22. Let $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ be two state sequences generated according to Equation (3.5) from α_0 and σ_0^2. Let ϵ be the constant required in Lemma 3.21. Then, the following holds for any $\delta > \delta_{\text{AMP}}$, $0 < \sigma_w^2 < \epsilon$, and $0 < \alpha_0 \leq 1$ and $\sigma_0^2 \leq 1$:

$$\lim_{t\to\infty} \alpha_t = \alpha_\star(\delta,\sigma_w^2) \quad \text{and} \quad \lim_{t\to\infty} \sigma_t^2 = \sigma_\star^2(\delta,\sigma_w^2),$$

where $\alpha_\star(\delta,\sigma_w^2)$ is the unique positive solution to $F_1^{-1}(\alpha)=F_2(\alpha;\delta,\sigma_w^2)$ and $\sigma_\star^2(\delta,\sigma_w^2)=F_1^{-1}(\alpha_\star(\delta,\sigma_w^2))$.

Proof. From Lemma 3.21-(a), when σ_w^2 is small enough, $(\alpha_t,\sigma_t^2)\in\mathcal{R}$ for all $t\geq 1$, where $\mathcal{R}\triangleq\{(\alpha,\sigma^2)|0<\alpha\leq 1,\,0\leq\sigma^2\leq\tilde{\sigma}_{\max}^2\}$, where $\tilde{\sigma}_{\max}^2=\max\{1,4/\delta\}+4\sigma_w^2$. We divide $\mathcal{R}$ into several regions and discuss the dynamical behaviors of the state evolution for different regions separately. Specifically, we define

$$\begin{aligned}
\mathcal{R}_0&\triangleq\left\{(\alpha,\sigma^2)|0<\alpha\leq 1,\ \pi^2/16<\sigma^2\leq\tilde{\sigma}_{\max}^2\right\},\\
\mathcal{R}_1&\triangleq\left\{(\alpha,\sigma^2)|F_1^{-1}(\alpha_*)\leq\sigma^2\leq\pi^2/16,\ F_1(\sigma^2)\leq\alpha\leq 1\right\},\\
\mathcal{R}_2&\triangleq\left\{(\alpha,\sigma^2)|0<\alpha\leq\alpha_*,\ 0\leq\sigma^2<F_1^{-1}(\alpha)\right\},\\
\mathcal{R}_3&\triangleq\left\{(\alpha,\sigma^2)|\alpha_*\leq\alpha\leq 1,\ 0\leq\sigma^2<F_1^{-1}(\alpha_*)\right\},
\end{aligned}\tag{3.64}$$

where $\alpha_*\approx 0.53$ was defined in Equation (3.24). Notice that $\alpha_\star(\delta,0)=1$, and therefore it is guaranteed that $\alpha_\star(\delta,\sigma_w^2)>\alpha_*$ for small enough σ_w^2. See Fig. 3.8 for illustration. To prove the lemma, we will prove the following arguments:

(i) If $(\alpha_{t_0},\sigma_{t_0}^2)\in\mathcal{R}_0$, then there exists a finite $T_1\geq 1$ such that $(\alpha_{t_0+T_1},\sigma_{t_0+T_1}^2)\in\mathcal{R}\backslash\mathcal{R}_0$.

(ii) If $(\alpha_{t_0},\sigma_{t_0}^2)\in\mathcal{R}_1\cup\mathcal{R}_2$ for $t_0\geq 1$ (i.e., after one iteration), then there exists a finite $T_2\geq 1$ such that $(\alpha_{t_0+T_2},\sigma_{t_0+T_2}^2)\in\mathcal{R}_3$.

(iii) We show that if $(\alpha_{t_0},\sigma_{t_0}^2)\in\mathcal{R}_3$ for $t_0\geq 0$, then $(\alpha_t,\sigma_t^2)\in\mathcal{R}_3$ for all $t>t_0$, and (α_t,σ_t^2) converges to $(\alpha_\star,\sigma_\star^2)$.

The proof of (i) is similar to that of Lemma 3.8 and therefore omitted here.

Proof of (ii): Following the proof of Lemma 3.7, we argue that if $(\alpha_t,\sigma_t^2)\in\mathcal{R}_1\cup\mathcal{R}_2$

then the following holds

$$\alpha_{t+1} \geq B_1(\alpha_t, \sigma_t^2) \quad \text{and} \quad \sigma_{t+1}^2 \geq B_2(\alpha_t, \sigma_t^2), \tag{3.65}$$

where $B_1(\alpha_t, \sigma_t^2) = \min\{\alpha_t, F_1(\sigma_t^2)\}$ and $B_2(\alpha_t, \sigma_t^2) = \max\left\{\sigma_t^2, F_1^{-1}(\alpha_t)\right\}$. Then, it is easy to show that $(\alpha_{t+1}, \sigma_{t+1}^2) \in \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$. Applying this recursively, we see that (α, σ^2) either moves to $\mathcal{R}_3$ at a certain time or stays in $\mathcal{R}_1 \cup \mathcal{R}_2$. We next prove that the latter case cannot happen. Suppose that $(\alpha_t, \sigma_t^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$ for $t \geq t_0$. If this is the case, then it can be shown that

$$B_1(\alpha_t, \sigma_t^2) \leq B_1(\alpha_{t+1}, \sigma_{t+1}^2) \quad \text{and} \quad B_2(\alpha_t, \sigma_t^2) \geq B_2(\alpha_{t+1}, \sigma_{t+1}^2), \quad \forall t > t_0. \tag{3.66}$$

On the other hand, since we assume $(\alpha_t, \sigma_t^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$ for $t \geq t_0$, B_1 is upper bounded by α_* and B_2 lower bounded by $F_1^{-1}(\alpha_*)$. Hence, this means the sequences B_1 and B_2 converges to α_* and $F_1^{-1}(\alpha_*)$, respectively. This cannot happen since there is no fixed point in $\mathcal{R}_1 \cup \mathcal{R}_2$.

The proof for Equation (3.65) and Equation (3.66) are basically the same as those for the noiseless counterparts and hence skipped here. Please refer to the proof of Lemma 3.7. We only need to show that some of the key inequalities used in the proof of Lemma 3.7 still hold in the noisy case, which have been listed in Lemma 3.21 (e) and (f).

Proof of (iii): Lemma 3.21-(c), (d) and (g) imply that $F_2 < F_1^{-1}(\alpha_*)$ for all $\alpha \in [\alpha_*, 1]$. Then, based on the strong global attractiveness of F_1 and F_2, it is easy to show that if $(\alpha_{t_0}, \sigma_{t_0}^2) \in \mathcal{R}_3$ then $(\alpha_t, \sigma_t^2) \in \mathcal{R}_3$ for all $t \geq t_0$. We have proved in Lemma 3.21-(d) that F_2 is a decreasing function of α on $[0, \hat{\alpha}]$ and increasing on $[\hat{\alpha}, 1]$, where $\alpha_\star < \hat{\alpha} < 1$. Then, the maximum of F_2 on $[\alpha_\star, 1]$ can only happen at

either $\alpha_\star$ or 1. We assume that the latter case happens; it will be clear that our proof for the former case is a special case of the proof for the latter one. See the right panel of Fig. 3.8.

As discussed above, we assume that $F_2(1;\delta,\sigma_w^2) > F_2(\alpha_\star;\delta,\sigma_w^2)$. Hence, by Lemma 3.21-(d), there exists a unique number $\alpha_\diamond \in (\alpha_\star, 1)$ such that $F_2(\alpha_\diamond;\delta,\sigma_w^2) = F_2(\alpha_\star;\delta,\sigma_w^2)$. See the plot in the right panel of Fig. 3.8. We further divide $\mathcal{R}_3$ into four regions:

$$\begin{aligned}
\mathcal{R}_{3a} &\triangleq \left\{(\alpha,\sigma^2)|\alpha_* \le \alpha \le \alpha_\diamond,\ F_1^{-1}(\alpha_\diamond) < \sigma^2 \le F_1^{-1}(\alpha_*)\right\},\\
\mathcal{R}_{3b} &\triangleq \left\{(\alpha,\sigma^2)|\alpha_* \le \alpha \le 1,\ 0 \le \sigma^2 < F_1^{-1}(\alpha_\diamond)\right\},\\
\mathcal{R}_{3c} &\triangleq \left\{(\alpha,\sigma^2)|\alpha_\diamond < \alpha \le 1,\ F_1^{-1}(\alpha_\diamond) \le \sigma^2 < F_1^{-1}(\alpha_*)\right\}.
\end{aligned}$$

Based on the strong global attractiveness of F_1 and F_2 (and similar to the proof of part (i) of this lemma), we can show the following:

- if $(\alpha_{t_0},\sigma_{t_0}^2) \in \mathcal{R}_{3a}$, then $(\alpha_{t_0+1},\sigma_{t_0+1}^2)$ can only be in $\mathcal{R}_{3a}$;
- if $(\alpha_{t_0},\sigma_{t_0}^2) \in \mathcal{R}_{3b}$, then $(\alpha_{t_0+1},\sigma_{t_0+1}^2)$ can be in $\mathcal{R}_{3a}$, $\mathcal{R}_{3b}$ or $\mathcal{R}_{3c}$;
- if $(\alpha_{t_0},\sigma_{t_0}^2) \in \mathcal{R}_{3c}$, then $(\alpha_{t_0+1},\sigma_{t_0+1}^2)$ can be in $\mathcal{R}_{3c}$ or $\mathcal{R}_{3a}$.

Putting things together, and similar to the treatment of $\mathcal{R}_0$, it can be shown that there exists a finite T_3 such that $(\alpha_t,\sigma_t^2) \in \mathcal{R}_{3a}$ for all $t \ge t_0 + T_3$.

It only remains to prove that if $(\alpha_{t'},\sigma_{t'}^2) \in \mathcal{R}_{3a}$ at a certain $t' \ge 0$, then $\{(\alpha_t,\sigma_t^2)\}_{t\ge t'}$

converges to $(\alpha_\star, \sigma_\star^2)$. To this end, define

$$
\begin{aligned}
B_1^{\text{low}}(\alpha, \sigma^2) &\triangleq \min\left\{\alpha_\star, \alpha, F_1(\sigma^2)\right\}, \\
B_1^{\text{up}}(\alpha, \sigma^2) &\triangleq \max\left\{\alpha_\star, \alpha, F_1(\sigma^2)\right\}, \\
B_2^{\text{low}}(\alpha, \sigma^2) &\triangleq \min\left\{\sigma_\star^2, \sigma^2, F_1^{-1}(\alpha)\right\} = F_1^{-1}\left(B_1^{\text{up}}(\alpha, \sigma^2)\right), \\
B_2^{\text{up}}(\alpha, \sigma^2) &\triangleq \max\left\{\sigma_\star^2, \sigma^2, F_1^{-1}(\alpha)\right\} = F_1^{-1}\left(B_1^{\text{low}}(\alpha, \sigma^2)\right).
\end{aligned}
$$

See examples depicted in Fig. 3.8. Using the strong global attractiveness of F_1 and F_2 and noting that $F_1^{-1}(\alpha) > F_2(\alpha) > \sigma_\star^2$ for $\alpha \in [\alpha_*, \alpha_\star)$ and $F_1^{-1}(\alpha) < F_2(\alpha) < \sigma_\star^2$ for $\alpha \in (\alpha_\star, \alpha_\diamond)$, it can be proved that

$$
\begin{aligned}
B_1^{\text{low}}(\alpha_t, \sigma_t^2) \le \alpha_{t+1} \le B_1^{\text{up}}(\alpha_t, \sigma_t^2), \\
B_2^{\text{low}}(\alpha_t, \sigma_t^2) \le \sigma_{t+1}^2 \le B_2^{\text{up}}(\alpha_t, \sigma_t^2).
\end{aligned}
$$

Further, the sequences $\{B_1^{\text{low}}(\alpha_t, \sigma_t^2)\}_{t \ge t'}$ and $\{B_2^{\text{low}}(\alpha_t, \sigma_t^2)\}_{t \ge t'}$ are monotonically non-decreasing and $\{B_1^{\text{up}}(\alpha_t, \sigma_t^2)\}_{t \ge t'}$ and $\{B_2^{\text{up}}(\alpha_t, \sigma_t^2)\}_{t \ge t'}$ are monotonically non-increasing. Also, B_1^{low} and B_2^{low} are upper bounded by $\alpha_\star$ and $\sigma_\star^2$, and B_1^{up} and B_2^{up} are lowered bounded by $\alpha_\star$ and $\sigma_\star^2$. Together with some arguments about the strict monotonicity of $\{B_1^{\text{low}}(\alpha_t, \sigma_t^2)\}_{t \ge t'}$ and $\{B_2^{\text{low}}(\alpha_t, \sigma_t^2)\}_{t \ge t'}$ (see discussions below Equation (3.50)), we have

$$
\begin{aligned}
\lim_{t\to\infty} B_1^{\text{low}}(\alpha_t, \sigma_t^2) = \lim_{t\to\infty} B_1^{\text{up}}(\alpha_t, \sigma_t^2) = \alpha_\star, \\
\lim_{t\to\infty} B_2^{\text{low}}(\alpha_t, \sigma_t^2) = \lim_{t\to\infty} B_2^{\text{up}}(\alpha_t, \sigma_t^2) = \sigma_\star^2,
\end{aligned}
$$

which implies that $\lim_{t\to\infty} \alpha_{t+1} = \alpha_\star$ and $\lim_{t\to\infty} \sigma_{t+1}^2 = \sigma_\star^2$. We skip the proofs for the above statements since similar arguments have been repeatedly used in this chapter. □

3.5.4 Proof of Theorem 3.4

According to Lemma 3.22, we know that (α_t, σ_t^2) converges to the unique fixed point of the state evolution equation. We now analyze the location of this fixed point and further derive the noise sensitivity. Applying a variable change $s \triangleq \sigma/\alpha$, we obtain the following equations for this unique fixed point:

$$\alpha = \phi_1(s), \tag{3.67a}$$

$$\sigma^2 = \frac{4}{\delta}\left\{\alpha^2 + \sigma^2 + 1 - \alpha\left[\phi_1(s) + \phi_3(s)\right]\right\} + 4\sigma_w^2, \tag{3.67b}$$

where ϕ_1 and ϕ_3 are defined in Equation (3.62). Using Equation (3.67a) and $\sigma^2 = \alpha^2 s^2 = \phi_1^2(s)s^2$, and after some algebra, we can write Equation (3.67b) as

$$T(s^2, \sigma_w^2) \triangleq \left(1 - \frac{4}{\delta}\right)\phi_1^2(s)s^2 + \frac{4}{\delta}\phi_1(s)\phi_3(s) - \left(\frac{4}{\delta} + 4\sigma_w^2\right) = 0. \tag{3.68}$$

Differentiating with respect to s^2 yields

$$\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2} = \left(1 - \frac{4}{\delta}\right)\left(\phi_1^2(s) + 2\phi_1(s)\frac{\mathrm{d}\phi_1(s)}{\mathrm{d}s^2}s^2\right) + \frac{4}{\delta}\frac{\mathrm{d}\phi_1(s)\phi_3(s)}{\mathrm{d}s^2}. \tag{3.69}$$

Using the identities listed in Equation (3.63), we have

$$\left.\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2}\right|_{s=0} = 1 - \frac{2}{\delta}.$$

Also, it is straightforward to see that $\frac{\partial T(s^2, \sigma_w^2)}{\partial \sigma_w^2} = -4$. Note that we have an implicit relation between s^2 and σ_w^2, and by the implicit function theorem we have

$$\lim_{\sigma_w^2 \to 0}\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2} = -\lim_{s^2 \to 0}\left(\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2}\right)^{-1}\frac{\partial T(s^2, \sigma_w^2)}{\partial \sigma_w^2} = \frac{4}{1 - \frac{2}{\delta}}.$$

Further, s is a continuously differentiable function of σ_w^2. Hence, by the mean value theorem we know that

$$\frac{s^2}{\sigma_w^2} = \left.\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2}\right|_{\tilde{\sigma}_w^2},$$

where $0 \leq \tilde{\sigma}_w \leq \sigma_w$. By taking $\lim_{\sigma_w^2 \to 0}$ from both sides of the above equality we have

$$\lim_{\sigma_w^2 \to 0} \frac{s^2}{\sigma_w^2} = \lim_{\tilde{\sigma}_w \to 0} \left.\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2}\right|_{\tilde{\sigma}_w^2} = -\lim_{s^2 \to 0} \left(\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2}\right)^{-1} \frac{\partial T(s^2, \sigma_w^2)}{\partial \sigma_w^2} = \frac{4}{1 - \frac{2}{\delta}}.$$

To derive the noise sensitivity, we notice that

$$\begin{aligned}\mathrm{AMSE}(\sigma_w^2, \delta) &= (\alpha - 1)^2 + \sigma^2 \\ &= [\phi_1(s) - 1]^2 + s^2 \phi_1^2(s).\end{aligned}$$

As shown in Equation (3.63), $\phi_1(s)$ can be expressed using elliptic integrals as:

$$\phi_1(s) = \sqrt{1 + s^2} E\left(\frac{1}{1 + s^2}\right) - \frac{s^2}{\sqrt{1 + s^2}} K\left(\frac{1}{1 + s^2}\right).$$

From Lemma 3.2-(i), $E(1 - \epsilon) = 1 + O(\epsilon \log \epsilon^{-1})$, hence $\sqrt{1 + s^2} E\left(\frac{1}{1+s^2}\right) = 1 + O(s^2 \log s^{-1})$. Further, since $K(1 - \epsilon) = O(\log \epsilon^{-1})$, we have $\frac{s^2}{\sqrt{1+s^2}} K\left(\frac{1}{1+s^2}\right) = O(s^2 \log s^{-1})$. Therefore, $\phi_1(s) - 1 = O(s^2 \log s^{-1})$. Hence, $\lim_{s^2 \to 0} \frac{[\phi_1(s)-1]^2}{s^2} = 0$ and so

$$\lim_{s^2 \to 0} \frac{\mathrm{AMSE}(\sigma_w^2, \delta)}{s^2} = \lim_{s^2 \to 0} \frac{[\phi_1(s) - 1]^2}{s^2} + \phi_1^2(s) = 1.$$

Finally,

$$\begin{aligned}\lim_{\sigma_w^2 \to 0} \frac{\mathrm{AMSE}(\sigma_w^2, \delta)}{\sigma_w^2} &= \lim_{s^2 \to 0} \frac{\mathrm{AMSE}(\sigma_w^2, \delta)}{s^2} \cdot \lim_{\sigma_w^2 \to 0} \frac{s^2}{\sigma_w^2} \\ &= \frac{4}{1 - \frac{2}{\delta}}.\end{aligned}$$

Chapter 4

Discussion and Conclusion

In this thesis, we have analyzed two non-convex optimization problems and two popular iterative algorithms that aim to solve those two optimizations. In Chapter 2, we provide global convergence guarantees of the EM algorithm for some specific mixture of two Gaussian models. Further, we have reveal some interesting properties of the EM algorithm in Section 2.4.2. Finally, we have provided a global analysis of the landscape of the maximum likelihood estimation of these models in Section 2.6. As a byproduct, our analysis on Model 3 and Model 4 provides some interesting insights to the mechanism of the over-parameterization techniques. In Chapter 3, we provide global analysis for the AMP algorithm on the phase retrieval problem for both complex valued signal and real valued signal. We have characterized the exact phase transition from global convergence to local convergence (i.e., Theorem 3.2 and Theorem 3.5) and from local convergence to when the algorithm has no chance to recover unless it is initialized at the correct solution (i.e., Theorem 3.3 and Theorem 3.6). We also analyze the noise sensitivity of the algorithm in the low noise regime (i.e., Theorem 3.4 and Theorem 3.7), and briefly discussed the impact of the ℓ_2 regularization and spectral initialization. Yet besides our results, there are still many

interesting questions for these two problems that remain unanswered. We here list some potential interesting future directions.

- **Extension to variance estimation for mixtures of two Gaussians:** Our current analysis only focus on mean/weight estimations and we assumed known variance. Naturally, our next step is to include variance estimation. The first promising model is the following:

  Model 5. The observation $\mathcal{Y}$ is an i.i.d. sample from the mixture distribution $0.5\mathcal{N}(-\boldsymbol{\theta}^\star, (\sigma^\star)^2\boldsymbol{I}) + 0.5\mathcal{N}(\boldsymbol{\theta}^\star, (\sigma^\star)^2\boldsymbol{I})$; $\sigma^\star$ is an unknown scalar, and $\boldsymbol{\theta}^\star$ is the unknown parameter of interest.

  We can show that the nice properties shown in Section 2.4.2 hold for Model 5 as well. Further the angles $\beta^{\langle t\rangle}$ between the estimate $\boldsymbol{\theta}^{\langle t\rangle}$ and $\boldsymbol{\theta}^\star$ are decreasing when $\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^\star\rangle \neq 0$. Finally, in one dimensional case, given $\theta^{\langle t\rangle}, \sigma^{\langle t\rangle} > 0$, we can show that not only the update rule of $\theta^{\langle t\rangle}$ has a unique fixed point for any given $\sigma^{\langle t\rangle}$, but also the update rule of $\sigma^{\langle t\rangle}$ has a unique fixed point for any given $\theta^{\langle t\rangle}$ as well. Further, we illustrate the fixed point curves in Figure 4.1. Although the relative positions do not satisfy the conditions given in Lemma 2.14, they are similar to the case of the noisy phase retrieval problem we have analyzed in Section 3.5 (See Figure 3.8). These benign properties suggest that we may have global convergence of $(\boldsymbol{\theta}^{\langle t\rangle}, \sigma^{\langle t\rangle})$ to the desired fixed points $(-\boldsymbol{\theta}^\star, \sigma^\star)$ or $(\boldsymbol{\theta}^\star, \sigma^\star)$ as long as $\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^\star\rangle \neq 0$.

  However, there is one main issue that makes our claim inconclusive. That is whether the EM algorithm can escape an undesired fixed point $(\boldsymbol{\theta}, \sigma^2) = (0, 1 + \frac{\|\boldsymbol{\theta}^\star\|^2}{d})$. More specifically, the Jacobian matrix with respect to $\boldsymbol{\theta}$ has $d-1$ number of small eigenvalues (specifically their values are $d/(d + \|\boldsymbol{\theta}^\star\|^2)$)

at $(\boldsymbol{\theta}, \sigma^2) = (0, 1 + \frac{\|\boldsymbol{\theta}^\star\|^2}{d})$. To address this issue, we may modified the EM algorithm to ensure a uniform lower bound on $\|\boldsymbol{\theta}^{\langle t \rangle}\|$ to avoid this undesired fixed point. Besides this issue, we are also missing some technical details to confirm the illustration of Figure 4.1 to complete the proof.

- **Optimal loss function for phase retrieval problem:** In this thesis, we mainly analyzed convergence of AMP algorithm for amplitude flow loss function. We have empirically evaluated Wirtinger flow loss function which has less performance than amplitude flow loss function. Further, we are able to show that it achieves the best convergence speed around the global optimum for the AMP algorithm when the loss function is indeed the amplitude loss. However, it is not clear whether our loss function gives the optimal landscape in the sense that the AMP algorithm achieves the optimal sample size to achieve convergence to the global optimum (either globally or under the best initialization scheme). Note that if we can answer this question, it may shed light on convergence of other algorithms and better understanding of the impact of loss function not only on phase retrieval but also other low rank matrix estimation as well.

Besides above possible future directions, one can also consider to follow our dynamical analysis to prove convergence for other non-convex optimization problems and iterative algorithms. Indeed, one of our main contributions is that our convergence analysis provides a better understanding towards the dynamics of the algorithms. As shown in the proof, our convergence analysis is carried out by understanding the directions of the movements from iteration t to iteration $t+1$ in every axis. We obtain these directional information by analyzing the fixed points of the update functions. We have shown that on each axis, under very mild conditions (e.g. monotonicity), the estimates should move towards the fixed point of the update function of that

axis. Therefore, by analyzing the relative position among the fixed points of different update functions, we may determine from which region the estimates will eventually escape (e.g. $\mathcal{R}_0$ in Figure 3.5) and in which region the estimates are remain trapped (e.g. $\mathcal{R}_1 \bigcup \mathcal{R}_2$ in Figure 3.5). Because of these dynamical understandings, we can conclude whether the estimates converge to the desired fixed point of the dynamical system. Furthermore, the transition of the relative positions of the fixed points may help us determine the transition between global convergence to local convergence (e.g. Theorem 3.2 and Theorem 3.3 in Chapter 3, and see Figure 3.4 for illustration).

- **Generalization of our geometric approach:** Despite of the success mentioned above, there is a major limitation of our current methods that is our target dynamical system has dimensions at most two. Therefore, we need to reduce the original high dimensional problem into a two dimensional one. In this thesis, this reduction is done by exploiting the special properties of both the problem and the algorithm (i.e., rotation invariant property of EM for GMM and state evolution from AMP framework for phase retrieval) which may not exist in general. Hence, how to generalize our methods to dynamical system with higher dimension becomes an interesting problem. On one hand, as discussed in Remark 2.4, our strategy is a successful generalization to the standard technique used in one dimension. It is possible to extend our approach further to three dimensions or higher. On the other hand, as dimension grows, the fixed point curves become hyperplane and the possible relative positions among them grow exponentially with dimensions. This could be the fundamental limits of our approach.

In summary, our results have shed some light on the convergence behavior of the iterative algorithms that target the non-convex optimization. We also hope that by

continuing our work on above interesting directions can provide deeper understanding of the dynamics of the algorithms in the future.

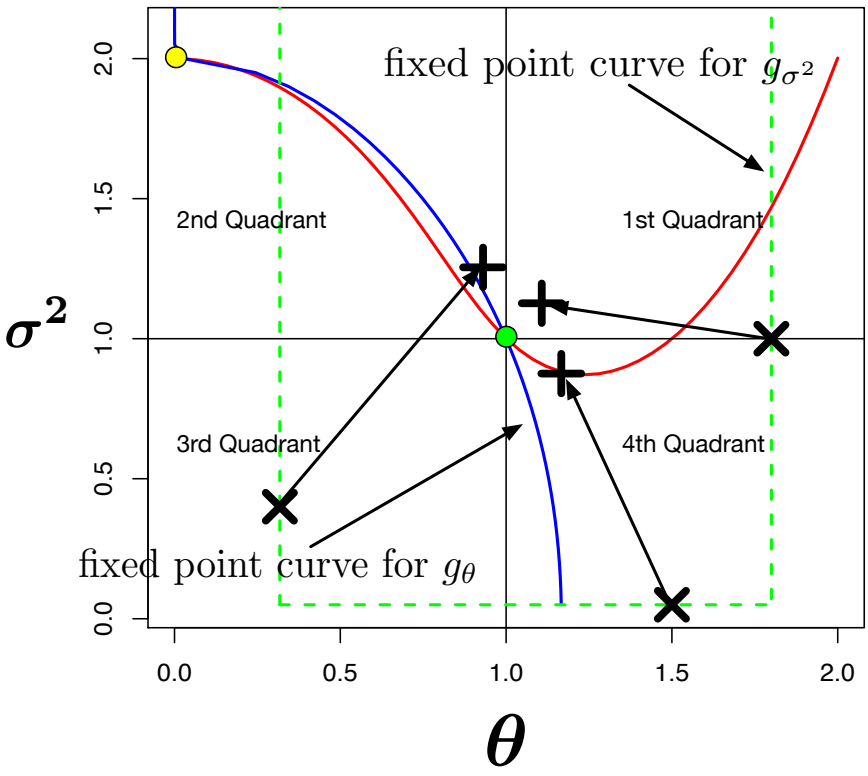

Figure 4.1: Let g_θ and g_{σ^2} be the update function for θ and σ^2 respectively. The fixed point curves for functions g_θ and g_w are shown with red and blue lines respectively. The green point at the intersections of the two curves is the correct convergence point $(\theta^\star, (\sigma^\star)^2) = (1, 1)$. The yellow point at the top left corner is the undesired fixed point $(0, 1 + (\theta^\star)^2)$. The cross points $\times$ are the possible initializations and the plus points $+$ are the corresponding positions after the first iteration. Let us divide the plane into 4 quadrants based on point $(\theta^\star, (\sigma^\star)^2)$. Initializations in 1st/3rd/4th quadrants will either converges to $(\theta^\star, (\sigma^\star)^2)$ or go to the 2nd quadrant. Initializations in the 2nd quadrant will be trapped in this quadrant and the relative positions of the fixed point curves is similar to Model 4 of EM.

Bibliography

E. Abbasi, F. Salehi, and B. Hassibi. Performance of real phase retrieval. In *International Conference on Sampling Theory and Applications (SampTA)*, July 2017.

D. Achlioptas and F. McSherry. On spectral learning of mixtures of distributions. In *Eighteenth Annual Conference on Learning Theory*, pages 458–469, 2005.

G. D. Anderson and M. K. Vamanamurthy. Inequalities for elliptic integrals. *Publ. Inst. Math.(Beograd)(NS)*, 37(51):61–63, 1985.

S. Arora and R. Kannan. Learning mixtures of separated nonspherical Gaussians. *The Annals of Applied Probability*, 15(1A):69–92, 2005.

S. Bahmani and J. Romberg. Phase retrieval meets statistical learning theory: A flexible convex relaxation. *arXiv preprint arXiv:1610.04210*, 2016.

M. Bakhshizadeh, A. Maleki, and S. Jalali. Compressive phase retrieval of structured signal. *arXiv preprint arXiv:1712.03278*, 2017.

S. Balakrishnan, M. J. Wainwright, and B. Yu. Statistical guarantees for the em algorithm: From population to sample-based analysis. *The Annals of Statistics*, 45(1):77–120, 02 2017.

R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, 2006.

A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A Nelson. Saving phase: injectivity and stability for phase retrieval. *Applied and Computational Harmonic Analysis*, 37(1):106–125, 2014.

J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová. Phase transitions, optimal errors and optimality of message-passing in generalized linear models. *arXiv preprint arXiv:1708.03395*, 2017.

M. Bayati and A. Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, Feb 2011.

M. Bayati and A. Montanari. The LASSO risk for Gaussian matrices. *IEEE Transactions on Information Theory*, 58(4):1997–2017, 2012.

M. Belkin and K. Sinha. Polynomial learning of distribution families. In *Fifty-First Annual IEEE Symposium on Foundations of Computer Science*, pages 103–112, 2010.

S. C. Brubaker and S. Vempala. Isotropic PCA and affine-invariant clustering. In *Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science*, 2008.

P. F. Byrd and M. D. Friedman. Handbook of elliptic integrals for engineers and scientists. 1971. *Berlin, Heidelberg. New York.*

T. Cai, X. Li, and Z. Ma. Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow. *The Annals of Statistics*, 44(5):2221–2251, 2016.

E. J. Candès and X. Li. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Foundations of Computational Mathematics*, 14(5):1017–1026, 2014.

E. J Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, April 2015.

K. Chaudhuri and S. Rao. Learning mixtures of product distributions using correlations and independence. In *Twenty-First Annual Conference on Learning Theory*, pages 9–20, 2008.

K. Chaudhuri, S. Dasgupta, and A. Vattani. Learning mixtures of gaussians using the k-means algorithm. *CoRR*, abs/0912.0086, 2009.

K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan. Multi-view clustering via canonical correlation analysis. In *Proceedings of the 26th annual international conference on machine learning*, 2009.

Y. Chen and E. J. Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Communications on Pure and Applied Mathematics*, 70:822–883, May 2017.

Y. Chen and Y. Chi. Harnessing structures in big data via guaranteed low-rank matrix estimation. *arXiv preprint arXiv:1802.08397*, 2018.

Y. Chi and Y. M. Lu. Kaczmarz method for solving quadratic equations. *IEEE Signal Processing Letters*, 23(9):1183–1187, 2016.

Y. Chi, Y. M. Lu, and Y. Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *arXiv preprint arXiv:1809.09573*, 2018.

S. Chrétien and A. O. Hero. On em algorithms and their proximal generalizations. *ESAIM: Probability and Statistics*, 12:308–326, May 2008.

D. Conniffe. Expected maximum log likelihood estimation. *Journal of the Royal Statistical Society. Series D*, 36(4):317–329, 1987.

S. Dasgupta and L. Schulman. A probabilistic analysis of EM for mixtures of separated, spherical Gaussians. *Journal of Machine Learning Research*, 8(Feb):203–226, 2007.

S. Dasgupta. Learning mixutres of Gaussians. In *Fortieth Annual IEEE Symposium on Foundations of Computer Science*, pages 634–644, 1999.

C. Daskalakis, C. Tzamos, and M. Zampetakis. Ten steps of em suffice for mixtures of two gaussians. In *Proceedings of the 2017 Conference on Learning Theory*, pages 704–710, 2017.

M. A. Davenport and J. Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016.

D. Davis, D. Drusvyatskiy, and C. Paquette. The nonsmooth landscape of phase retrieval. *arXiv preprint arXiv:1711.03247*, 2017.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1–38, 1977.

O. Dhifallah and Y. M. Lu. Fundamental limits of PhaseMax for phase retrieval: A replica analysis. *arXiv preprint arXiv:1708.03355*, 2017.

O. Dhifallah, C. Thrampoulidis, and Y. M. Lu. Phase retrieval via linear programming: Fundamental limits and algorithmic improvements. *arXiv preprint arXiv:1710.05234*, 2017.

D. L. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

J. C. Duchi and F. Ruan. Solving (most) of a set of quadratic equalities: composite optimization for robust phase retrieval. *arXiv preprint arXiv:1705.02356*, 2017.

Y. C. Eldar and S. Mendelson. Phase retrieval: Stability and recovery guarantees. *Applied and Computational Harmonic Analysis*, 36(3):473–494, 2014.

R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, London, A.*, 222:309–368, 1922.

T. Goldstein and C. Studer. PhaseMax: Convex phase retrieval via basis pursuit. *arXiv preprint arXiv:1610.07531*, 2016.

E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1-minimization: methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

M. Hardt and E. Price. Tight bounds for learning a mixture of two gaussians. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 753–760, 2015.

D. Hsu and S. M. Kakade. Learning mixtures of spherical Gaussians: moment methods and spectral decompositions. In *Fourth Innovations in Theoretical Computer Science*, 2013.

P. Jain and P. Kar. Non-convex optimization for machine learning. *Foundations and Trends® in Machine Learning*, 10(3-4):142–336, 2017.

S. Jalali and A. Maleki. From compression to compressed sensing. *Applied and Computational Harmonic Analysis*, 40(2):352–385, 2016.

A. Javanmard and A. Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115, 2013.

H. Jeong and C. S. Güntürk. Convergence of the randomized kaczmarz method for phase retrieval. *arXiv preprint arXiv:1706.10291*, 2017.

A. T. Kalai, A. Moitra, and G. Valiant. Efficiently learning mixtures of two Gaussians. In *Forty-second ACM Symposium on Theory of Computing*, pages 553–562, 2010.

R. Kannan, H. Salmasian, and S. Vempala. The spectral method for general mixture models. *SIAM Journal on Computing*, 38(3):1141–1156, 2008.

V. Koltchinskii. Oracle inequalities in empirical risk minimization and sparse recovery problems. In *École d'été de probabilités de Saint-Flour XXXVIII*, 2011.

F. C. Leone, L. S. Nelson, and R. B. Nottingham. The folded normal distribution. *Technometrics*, 3(4):543–550, 1961.

G. Li, Y. Gu, and Y. M. Lu. Phase retrieval using iterative projections: dynamics in the large systems limit. In *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1114–1118, Sept 2015.

S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions of Information Theory*, 28(2):129–137, 1982.

Y. M. Lu and G. Li. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. *arXiv preprint arXiv:1702.06435*, 2017.

C. Ma, K. Wang, Y. Chi, and Y. Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *arXiv preprint arXiv:1711.10467*, 2017.

J. Ma, J. Xu, and A. Maleki. Approximate message passing for amplitude based optimization. *arXiv preprint arXiv:1806.03276*, 2018.

J. Ma, J. Xu, and A. Maleki. Optimization-based amp for phase retrieval: The impact of initialization and ℓ_2 regularization. *IEEE Transactions on Information Theory*, 65(6):3600–3629, 2019.

J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

A. Moitra and G. Valiant. Settling the polynomial learnability of mixtures of Gaussians. In *Fifty-First Annual IEEE Symposium on Foundations of Computer Science*, pages 93–102, 2010.

M. Mondelli and A. Montanari. Fundamental limits of weak recovery with applications to phase retrieval. *arXiv preprint arXiv:1708.05932*, 2017.

A. Mousavi, A. Maleki, and R. G. Baraniuk. Consistent parameter estimation for LASSO and approximate message passing. *arXiv preprint arXiv:1511.01017*, 2015.

P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pages 2796–2804, 2013.

K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society, London, A.*, 185:71–110, 1894.

Q. Qu, Y. Zhang, Y. C. Eldar, and J. Wright. Convolutional phase retrieval via gradient descent. *arXiv preprint arXiv:1712.00716*, 2017.

S. Rangan. Generalized approximate message passing for estimation with random linear mixing. In *IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172, July 2011.

R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, 1984.

P. Schniter and S. Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2015.

Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

M. Soltanolkotabi. Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization. *arXiv preprint arXiv:1702.06175*, 2017.

J. Sun, Q. Qu, and J. Wright. A geometric analysis of phase retrieval. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2379–2383, July 2016.

Y. S. Tan and R. Vershynin. Phase retrieval via randomized kaczmarz: Theoretical guarantees. *arXiv preprint arXiv:1706.09993*, 2017.

C. Thrampoulidis, S. Oymak, and B. Hassibi. Regularized linear regression: A precise analysis of the estimation error. In *Conference on Learning Theory*, pages 1683–1709, 2015.

C. Thrampoulidis, E. Abbasi, and B. Hassibi. Precise error analysis of regularized m-estimators in high-dimensions. *arXiv preprint arXiv:1601.06233*, 2016.

P. Tseng. An analysis of the EM algorithm and entropy-like proximal point methods. *Mathematics of Operations Research*, 29(1):27–44, Feb 2004.

S. Vempala and G. Wang. A spectral algorithm for learning mixtures models. *Journal of Computer and System Sciences*, 68(4):841–860, 2004.

I. Waldspurger, A. d'Aspremont, and S. Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1-2):47–81, 2015.

M. K. Wang and Y. M. Chu. Asymptotical bounds for complete elliptic integrals of the second kind. *Journal of Mathematical Analysis and Applications*, 402(1):119–126, 2013.

G. Wang, G. B. Giannakis, and Y. C. Eldar. Solving systems of random quadratic equations via truncated amplitude flow. *arXiv preprint arXiv:1605.08285*, 2016.

K. Wei. Solving systems of phaseless equations via kaczmarz methods: A proof of concept study. *Inverse Problems*, 31(12):125008, 2015.

C. F. Jeff Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, Mar 1983.

L. Xu and M. I. Jordan. On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Computation*, 8:129–151, 1996.

J. Xu, D. J. Hsu, and A. Maleki. Global analysis of expectation maximization for mixtures of two gaussians. In *Advances in Neural Information Processing Systems*, pages 2676–2684, 2016.

J. Xu, D. J. Hsu, and A. Maleki. Benefits of over-parameterization with em. In *Advances in Neural Information Processing Systems*, pages 10662–10672, 2018.

W. J. Zeng and H. C. So. Coordinate descent algorithms for phase retrieval. *arXiv preprint arXiv:1706.03474*, 2017.

H. Zhang and Y. Liang. Reshaped wirtinger flow for solving quadratic system of equations. In *Advances in Neural Information Processing Systems*, pages 2622–2630, 2016.

L. Zheng, A. Maleki, H. Weng, X. Wang, and T. Long. Does ℓ_p-minimization outperform ℓ_1-minimization? *IEEE Transactions on Information Theory*, PP(99):1–1, 2017.

L. Zhu. On a quadratic estimate of shafer. *J. Math. Inequal*, 2(4):571–574, 2008.

Appendix A

Proofs omitted in main chapters

A.1 Proofs of Population EM results omitted in Section 2.4

A.1.1 Proofs omitted in Sections 2.4.1

A.1.1.1 Proof of Lemma 2.2

In the proof of this lemma, we have $h(\theta, w_1) = H(\theta; \theta^\star, w_1)$. To prove Equation (2.38), we just need to show

$$\frac{\partial h(\theta, w_1)}{\partial w_1} \begin{cases} > 0, & w_1 > 0.5 \\ < 0, & w_1 < 0.5 \end{cases} \quad \forall \theta < \theta^\star. \tag{A.1}$$

To prove this, we divide it into two cases (i) $\theta \leq 0$ and (ii) $\theta \in (0, \theta^\star)$. To prove (i), by the definition of $h(\theta, w_1)$ in Equation (2.24) (with $w_2 = 1 - w_1$), we have

$$\begin{aligned} \frac{\partial h(\theta, w_1)}{\partial w_1} &= \underbrace{\int \frac{w_1 e^{y\theta} - w_2 e^{-y\theta}}{w_1 e^{y\theta} + w_2 e^{-y\theta}} y(\phi(y - \theta^\star) - \phi(y + \theta^\star)) \mathrm{d}y}_{\text{part 1}} \\ &\quad + 2 \underbrace{\int \frac{w_1 e^{y\theta^\star} + w_2 e^{-y\theta^\star}}{(w_1 e^{y\theta} + w_2 e^{-y\theta})^2} y\phi(y) e^{-\frac{(\theta^\star)^2}{2}} \mathrm{d}y}_{\text{part 2}}. \end{aligned}$$

For part 1, we have

$$\begin{aligned}\text{part 1} &= \int_{y\geq 0}\left\{\frac{w_1e^{y\theta}-w_2e^{-y\theta}}{w_1e^{y\theta}+w_2e^{-y\theta}}+\frac{w_1e^{-y\theta}-w_2e^{y\theta}}{w_1e^{-y\theta}+w_2e^{y\theta}}\right\}y(\phi(y-\theta^\star)-\phi(y+\theta^\star))\mathrm{d}y\\ &= 2\int_{y\geq 0}\frac{w_1^2-w_2^2}{w_1^2+w_2^2+w_1w_2(e^{-y\theta}+e^{y\theta})}y(\phi(y-\theta^\star)-\phi(y+\theta^\star))\mathrm{d}y\end{aligned}$$

Hence, we have

$$\text{part 1}\begin{cases}>0, & w_1>0.5\\ <0, & w_1<0.5\end{cases}. \tag{A.2}$$

For part 2, we have

$$\begin{aligned}\text{part 2} &= \int_{y\geq 0}\left\{\frac{w_1e^{y\theta^\star}+w_2e^{-y\theta^\star}}{(w_1e^{y\theta}+w_2e^{-y\theta})^2}-\frac{w_1e^{-y\theta^\star}+w_2e^{y\theta^\star}}{(w_1e^{-y\theta}+w_2e^{y\theta})^2}\right\}y\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y\\ &= (w_1-w_2)\int_{y\geq 0}\left\{\frac{(w_1^2+w_2^2+w_1w_2)(e^{y(\theta^\star-2\theta)}-e^{y(2\theta-\theta^\star)})+2w_1w_2(e^{y\theta^\star}-e^{-y\theta^\star})}{(w_1e^{y\theta}+w_2e^{-y\theta})^2(w_1e^{-y\theta}+w_2e^{y\theta})^2}\right.\\ &\quad\left.+\frac{w_1w_2(e^{-y(\theta^\star+2\theta)}-e^{y(\theta^\star+2\theta)})}{(w_1e^{y\theta}+w_2e^{-y\theta})^2(w_1e^{-y\theta}+w_2e^{y\theta})^2}\right\}y\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y.\end{aligned}$$

Since $\theta\leq 0$, we have

$$e^{y(\theta^\star-2\theta)}-e^{y(2\theta-\theta^\star)} \geq \max\left\{\left|e^{y\theta^\star}-e^{-y\theta^\star}\right|,\left|e^{y(\theta^\star+2\theta)}-e^{-y(\theta^\star+2\theta)}\right|\right\}.$$

Hence, we have

$$\frac{\text{part 2}}{w_1-w_2} \geq \int_{y\geq 0}\frac{(w_1-w_2)^2(e^{y(\theta^\star-2\theta)}-e^{y(2\theta-\theta^\star)})}{(w_1e^{y\theta}+w_2e^{-y\theta})^2(w_1e^{-y\theta}+w_2e^{y\theta})^2}y\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y \geq 0.$$

Therefore, we have

$$\text{part 2}\begin{cases}\geq 0, & w_1>0.5\\ \leq 0, & w_1<0.5\end{cases}. \tag{A.3}$$

Combine Equation (A.2) and Equation (A.3), we have Equation (A.1) holds for case (i). To prove case (ii), we use a different strategy. First note that $h(\theta^\star,w)\equiv\theta^\star$, hence,

$$\left.\frac{\partial h(\theta,w)}{\partial w}\right|_{\theta=\theta^\star} = 0. \tag{A.4}$$

Therefore, to prove Equation (A.1) for case (ii), we just need to show

$$\frac{\partial^2 h(\theta, w_1)}{\partial\theta\partial w_1} \begin{cases} < 0, & w_1 > 0.5 \\ > 0, & w_1 < 0.5 \end{cases} \quad \forall\theta \in (0, \theta^\star). \tag{A.5}$$

By the definition of $h(\theta, w_1)$ in Equation (2.24) (with $w_2 = 1 - w_1$), we have

$$\begin{aligned} \frac{1}{4}\frac{\partial^2 h(\theta, w_1)}{\partial\theta\partial w_1} &= \underbrace{2w_1w_2\int \frac{e^{y(\theta^\star-\theta)} - e^{y(\theta-\theta^\star)}}{(w_1e^{y\theta} + w_2e^{-y\theta})^3} y^2\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y}_{\text{part 3}} \\ &\quad + \underbrace{\int \left(\frac{w_2^2e^{-y\theta^\star}}{(w_1e^{y\theta} + w_2e^{-y\theta})^2} - \frac{w_1^2e^{y\theta^\star}}{(w_1e^{y\theta} + w_2e^{-y\theta})^2}\right) y^2\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y}_{\text{part 4}} \end{aligned}$$

For part 3, we have

$$\begin{aligned} &\frac{\text{part 3}}{2w_1w_2} \\ &= \int_{y\geq 0} (w_1 - w_2)\frac{(e^{y(\theta^\star-\theta)} - e^{y(\theta-\theta^\star)})(e^{-y\theta} - e^{y\theta})(A^2 + B^2 - AB)}{(w_1e^{y\theta} + w_2e^{-y\theta})^3(w_1e^{-y\theta} + w_2e^{y\theta})^3} y^2\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y, \end{aligned}$$

where $A = w_1e^{y\theta} + w_2e^{-y\theta}$ and $B = w_1e^{-y\theta} + w_2e^{y\theta}$. Hence, since $\theta \in (0, \theta^\star)$, we have

$$\text{part 3} \begin{cases} < 0, & w_1 > 0.5 \\ > 0, & w_1 < 0.5 \end{cases}. \tag{A.6}$$

For part 4, we have

$$\begin{aligned} &\frac{\text{part 4}}{(w_1^2 + w_2^2)} \\ &= -\int_{y\geq 0} (w_1 - w_2)\frac{(e^{(2\theta-\theta^\star)y} + e^{-(2\theta-\theta^\star)y}) + 2w_1w_2(e^{y\theta^\star} + e^{-y\theta^\star})}{(w_1e^{y\theta} + w_2e^{-y\theta})^2(w_1e^{-y\theta} + w_2e^{y\theta})^2} y^2\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y. \end{aligned}$$

Hence, we have

$$\text{part 4} \begin{cases} < 0, & w_1 > 0.5 \\ > 0, & w_1 < 0.5 \end{cases}. \tag{A.7}$$

Combine Equation (A.6) and Equation (A.7), we have Equation (A.5) holds and therefore Equation (A.1) holds for case (ii). This completes the proof for Equation

(2.38). To prove Equation (2.39), note that

$$\begin{aligned} 0 \;\le\; \frac{\partial H(\theta, w_1)}{\partial \theta} &= \int \frac{4w_1w_2}{(w_1e^{y\theta}+w_2e^{-y\theta})^2} y^2(w_1\phi(y-\theta^\star)+w_2\phi(y+\theta^\star))\mathrm{d}y \\ &= \underbrace{\int_{y\ge 0} \frac{4w_1w_2}{(w_1e^{y\theta}+w_2e^{-y\theta})^2} y^2(w_1\phi(y-\theta^\star)+w_2\phi(y+\theta^\star))\mathrm{d}y}_{\text{part 5}} \\ &\quad + \underbrace{\int_{y\ge 0} \frac{4w_1w_2}{(w_2e^{y\theta}+w_1e^{-y\theta})^2} y^2(w_2\phi(y-\theta^\star)+w_1\phi(y+\theta^\star))\mathrm{d}y}_{\text{part 6}}. \end{aligned}$$

Since part 5 and part 6 are symmetric with respect to w_1, w_2, WLOG, we assume $w_1 \ge 0.5$. Then for part 5, note that since $\theta \ge \theta^\star$, we have $w_1e^{y\theta^\star}+w_2e^{-y\theta^\star} \le w_1e^{y\theta}+w_2e^{-y\theta}$, and therefore,

$$\begin{aligned} \text{part 5} \;&\le\; \int_{y\ge 0} \frac{4w_1w_2}{(w_1e^{y\theta^\star}+w_2e^{-y\theta^\star})^2} y^2(w_1\phi(y-\theta^\star)+w_2\phi(y+\theta^\star))\mathrm{d}y \\ &= \int_{y\ge 0} \frac{4w_1w_2}{w_1e^{y\theta^\star}+w_2e^{-y\theta^\star}} y^2\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y \\ &\le \int_{y\ge 0} 2\sqrt{w_1w_2}\, y^2\phi(y)e^{-\frac{(\theta^\star)^2}{2}}\mathrm{d}y \;\le\; \frac{e^{-\frac{(\theta^\star)^2}{2}}}{2}, \end{aligned} \tag{A.8}$$

where last two inequalities hold due to AM-GM inequality. For part 6, we have if $\theta \ge \theta^\star$,

$$\begin{aligned} \text{part 6} \;&=\; \int_{y\ge 0} \frac{4}{\left(\sqrt{\frac{w_1}{w_2}}e^{-y\theta}+\sqrt{\frac{w_2}{w_1}}e^{y\theta}\right)^2} y^2(w_1\phi(y+\theta^\star)+w_2\phi(y-\theta^\star))\mathrm{d}y \\ &\overset{\text{(i)}}{\le} \int_{y\ge 0} \frac{2}{e^{(y-\frac{\ln(w_1/w_2)}{2\theta})\theta}+e^{-(y-\frac{\ln(w_1/w_2)}{2\theta})\theta}} y^2(w_1\phi(y+\theta^\star)+w_2\phi(y-\theta^\star))\mathrm{d}y \\ &\overset{\text{(ii)}}{\le} \int_{y\ge 0} \frac{2}{e^{(y-\frac{\ln(w_1/w_2)}{2\theta})\theta^\star}+e^{-(y-\frac{\ln(w_1/w_2)}{2\theta})\theta^\star}} y^2(w_1\phi(y+\theta^\star)+w_2\phi(y-\theta^\star))\mathrm{d}y \\ &= \int_{y\ge 0} \frac{2}{(\frac{w_2}{w_1})^{\frac{\theta^\star}{2\theta}}e^{y\theta^\star}+(\frac{w_1}{w_2})^{\frac{\theta^\star}{2\theta}}e^{-y\theta^\star}} y^2(w_1\phi(y+\theta^\star)+w_2\phi(y-\theta^\star))\mathrm{d}y, \end{aligned} \tag{A.9}$$

where inequality (i) holds due to AM-GM inequality, and inequality (ii) holds due to the monotonic of hyperbolic cosine function. Our next step is to prove for all $y\theta^\star \ge 0$

and $0 < \theta^\star \leq \theta$, we have

$$(\frac{w_2}{w_1})^{\frac{\theta^\star}{2\theta}} e^{y\theta^\star} + (\frac{w_1}{w_2})^{\frac{\theta^\star}{2\theta}} e^{-y\theta^\star} \geq 2(w_1 e^{-y\theta^\star} + w_2 e^{y\theta^\star}), \tag{A.10}$$

which, with Equation (A.9), immediately implies that

$$\text{part 6} \;\leq\; \int_{y\geq 0} y^2 \phi(y) e^{-\frac{(\theta^\star)^2}{2}} \mathrm{d}y \;=\; \frac{e^{-\frac{(\theta^\star)^2}{2}}}{2},$$

and therefore, combine with Equation (A.8), we have Equation (2.39) holds. To prove Equation (A.10), note that this is equivalent to prove

$$\left(\frac{w_2}{w_1}\right)^{\frac{\theta^\star}{2\theta}} \left(1 - 2w_1^{\frac{\theta^\star}{2\theta}} w_2^{1-\frac{\theta^\star}{2\theta}}\right) e^{y\theta^\star} \geq \left(\frac{w_1}{w_2}\right)^{\frac{\theta^\star}{2\theta}} \left(2w_2^{\frac{\theta^\star}{2\theta}} w_1^{1-\frac{\theta^\star}{2\theta}} - 1\right) e^{-y\theta^\star}. \tag{A.11}$$

Note that

$$\begin{aligned} w_1^{\frac{\theta^\star}{2\theta}} w_2^{1-\frac{\theta^\star}{2\theta}} + w_1^{1-\frac{\theta^\star}{2\theta}} w_2^{\frac{\theta^\star}{2\theta}} &= (w_1 w_2)^{\frac{\theta^\star}{2\theta}} (w_1^{1-\frac{\theta^\star}{\theta}} + w_2^{1-\frac{\theta^\star}{\theta}}) \\ &\leq \frac{w_1^{1-\frac{\theta^\star}{\theta}} + w_2^{1-\frac{\theta^\star}{\theta}}}{2^{\frac{\theta^\star}{\theta}}} \\ &\leq (w_1 + w_2)^{1-\frac{\theta^\star}{\theta}} \;=\; 1. \end{aligned}$$

where the last two inequalities holds due to AM-GM inequality and Holder inequality respectively. Also, since $w_1 \geq w_2$, we have

$$w_1^{\frac{\theta^\star}{2\theta}} w_2^{1-\frac{\theta^\star}{2\theta}} \leq w_1^{1-\frac{\theta^\star}{2\theta}} w_2^{\frac{\theta^\star}{2\theta}}.$$

Hence, we have

$$1 - 2w_1^{\frac{\theta^\star}{2\theta}} w_2^{1-\frac{\theta^\star}{2\theta}} \;\geq\; 0.$$

Therefore, to prove Equation (A.11), it is sufficient to prove

$$\left(\frac{w_2}{w_1}\right)^{\frac{\theta^\star}{2\theta}} \left(1 - 2w_1^{\frac{\theta^\star}{2\theta}} w_2^{1-\frac{\theta^\star}{2\theta}}\right) \;\geq\; \left(\frac{w_1}{w_2}\right)^{\frac{\theta^\star}{2\theta}} \left(2w_2^{\frac{\theta^\star}{2\theta}} w_1^{1-\frac{\theta^\star}{2\theta}} - 1\right),$$

which is equivalent to

$$\left(\frac{w_2}{w_1}\right)^{\frac{\theta^\star}{2\theta}} + \left(\frac{w_1}{w_2}\right)^{\frac{\theta^\star}{2\theta}} \geq 2(w_1 + w_2) = 2,$$

which holds due to AM-GM inequality. Hence, we have Equation (A.10) holds.

A.1.2 Proofs omitted in Sections 2.4.2

A.1.2.1 Proof of Lemma 2.4

We first show Lemma 2.4. Recall that, after reparameterization, the Population EM estimates for Model 2 satisfy the update rule in Equation (2.15) and Equation (2.16), i.e.,

$$\begin{aligned}
\boldsymbol{a}^{\langle t+1\rangle} &= \frac{\boldsymbol{\gamma}^{\langle t+1\rangle}(1-2\mathsf{p}^{\langle t+1\rangle})}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})}, \\
\boldsymbol{\theta}^{\langle t+1\rangle} &= \frac{\boldsymbol{\gamma}^{\langle t+1\rangle}}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})},
\end{aligned}$$

where (with Equation (2.27) and Equation (2.28))

$$\begin{aligned}
\boldsymbol{\gamma}^{\langle t+1\rangle} &= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star}\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} \;=\; g_\gamma(\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle};\boldsymbol{\theta}^\star)\,, \\
\mathsf{p}^{\langle t+1\rangle} &= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star}\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}) \;=\; g_p(\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle};\boldsymbol{\theta}^\star)\,.
\end{aligned}$$

Note that

$$\begin{aligned}
g_\gamma(\boldsymbol{a},\boldsymbol{\theta};\boldsymbol{\theta}^\star) &= g_\gamma(-\boldsymbol{a},\boldsymbol{\theta};\boldsymbol{\theta}^\star) \;=\; -g_\gamma(\boldsymbol{a},-\boldsymbol{\theta};\boldsymbol{\theta}^\star), \\
g_p(\boldsymbol{a},\boldsymbol{\theta};\boldsymbol{\theta}^\star) &= 1-g_p(-\boldsymbol{a},\boldsymbol{\theta};\boldsymbol{\theta}^\star) \;=\; 1-g_p(\boldsymbol{a},-\boldsymbol{\theta};\boldsymbol{\theta}^\star).
\end{aligned}$$

With induction, it is straightforward to show the second half of the lemma and we just need to show that for all $t \geq 0$,

$$\begin{aligned}
\operatorname{sgn}\left(\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle\right) &= \operatorname{sgn}\left(\langle \boldsymbol{a}^{\langle t+1\rangle},\boldsymbol{\theta}^{\langle t+1\rangle}\rangle\right) \;=\; \operatorname{sgn}\left(\frac{1}{2}-\mathsf{p}^{\langle t+1\rangle}\right), \\
\operatorname{sgn}\left(\langle \boldsymbol{\theta}^{\langle t\rangle},\boldsymbol{\theta}^\star\rangle\right) &= \operatorname{sgn}\left(\langle \boldsymbol{\theta}^{\langle t+1\rangle},\boldsymbol{\theta}^\star\rangle\right).
\end{aligned}$$

From Equation (2.15) and Equation (2.16), we know that

$$\begin{aligned}\langle \boldsymbol{\theta}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle &= \frac{\langle \boldsymbol{\gamma}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle}{\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})},\\ \langle \boldsymbol{a}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\langle t+1\rangle}\rangle &= \|\boldsymbol{\theta}^{\langle t+1\rangle}\|^2(1-2\mathsf{p}^{\langle t+1\rangle}).\end{aligned}$$

Since $\mathsf{p}^{\langle t+1\rangle} \in (0,1)$ and $\|\boldsymbol{\theta}^{\star}\| > 0$, we have

$$\begin{aligned}\operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle) &= \operatorname{sgn}(\langle \boldsymbol{\gamma}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle)\\ \operatorname{sgn}(\langle \boldsymbol{a}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\langle t+1\rangle}\rangle) &= \operatorname{sgn}(1-2\mathsf{p}^{\langle t+1\rangle})\operatorname{sgn}(\|\boldsymbol{\theta}^{\langle t+1\rangle}\|^2)\\ &= \operatorname{sgn}(1-2\mathsf{p}^{\langle t+1\rangle})|\operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle)|.\end{aligned}$$

Hence if we have

$$\operatorname{sgn}(\langle \boldsymbol{\gamma}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle) = \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^{\star}\rangle), \tag{A.12}$$

and

$$\operatorname{sgn}(1-2\mathsf{p}^{\langle t+1\rangle}) = \operatorname{sgn}(\langle \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\rangle), \tag{A.13}$$

then immediately, we have

$$\operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle) = \operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^{\star}\rangle),$$

and

$$\begin{aligned}\operatorname{sgn}(\langle \boldsymbol{a}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\langle t+1\rangle}\rangle) &= \operatorname{sgn}(1-2\mathsf{p}^{\langle t+1\rangle})|\operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle t+1\rangle}, \boldsymbol{\theta}^{\star}\rangle)|\\ &= \operatorname{sgn}(\langle \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\rangle)|\operatorname{sgn}(\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^{\star}\rangle)|\\ &= \operatorname{sgn}(\langle \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\rangle).\end{aligned}$$

Hence our next goal is to prove Equation (A.12) and Equation (A.13). Note that for all orthogonal matrix $\boldsymbol{V} \in \mathbb{R}^{d\times d}$ and any vectors $\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, we have

$$\begin{aligned}g_{\gamma}(\boldsymbol{V}\boldsymbol{a}, \boldsymbol{V}\boldsymbol{b}; \boldsymbol{V}\boldsymbol{\theta}^{\star}) &= \boldsymbol{V} g_{\gamma}(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\theta}^{\star}),\\ g_{p}(\boldsymbol{V}\boldsymbol{a}, \boldsymbol{V}\boldsymbol{b}; \boldsymbol{V}\boldsymbol{\theta}^{\star}) &= g_{p}(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\theta}^{\star}).\end{aligned} \tag{A.14}$$

Hence, it is straightforward to check that all the quantities $\langle \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\rangle, \langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^{\star}\rangle, 1-2\mathsf{p}^{\langle t\rangle}$ are rotation invariant. Therefore, with Lemma 2.3, to show Equation (A.12)

and Equation (A.13), we can assume without loss of generality that $\boldsymbol{\theta}^\star$ has all its coordinates except the first one equal to zero, i.e., $\boldsymbol{\theta}^\star = (\|\boldsymbol{\theta}^\star\|, 0, \ldots, 0)^\top$. Then, we have

$$\begin{aligned}
&\langle \boldsymbol{\gamma}^{\langle t+1\rangle}, \boldsymbol{\theta}^\star\rangle \;=\; \langle g_\gamma(\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}; \boldsymbol{\theta}^\star), \boldsymbol{\theta}^\star\rangle \\
=\;& \|\boldsymbol{\theta}^\star\| \int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) y_1 \phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}^\star) \mathrm{d}\boldsymbol{y} \\
=\;& \|\boldsymbol{\theta}^\star\| \int \frac{e^{-\sum_{i=2}^d y_i^2/2}}{\sqrt{2\pi}^{d-1}} \Big(\int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) y_1 \phi^+(y_1, \|\boldsymbol{\theta}^\star\|) \mathrm{d}y_1 \Big) \mathrm{d}y_2 \cdots \mathrm{d}y_d,
\end{aligned} \tag{A.15}$$

where $\phi^+(y, \theta)$ is a shorthand for $\phi^+(y, \theta, 0.5)$ in this proof. Hence, define $\boldsymbol{y}_{2\ldots d} = (y_2, \cdots, y_d)$ and $B(\boldsymbol{y}_{2\ldots d}) \triangleq \sum_{i=2}^d y_i b_i^{\langle t\rangle}$, then to prove Equation (A.12), it is sufficient to prove

$$\begin{aligned}
\mathrm{sgn}\Big(\int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) y_1 \phi^+(y_1, \|\boldsymbol{\theta}^\star\|) \mathrm{d}y_1 \Big) \;&=\; \mathrm{sgn}(\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^\star\rangle) \;=\; \mathrm{sgn}(\|\boldsymbol{\theta}^\star\| \theta_1^{\langle t\rangle}) \\
&=\; \mathrm{sgn}(\theta_1^{\langle t\rangle}), \quad \forall \boldsymbol{y}_{2\ldots d}, B(\boldsymbol{y}_{2\ldots d}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) y_1 \phi^+(y_1, \|\boldsymbol{\theta}^\star\|) \mathrm{d}y_1 \\
=\;& \int_0^{+\infty} \big(\mathsf{w}_d((y_1, \boldsymbol{y}_{2\ldots d})^\top - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) - \mathsf{w}_d((-y_1, \boldsymbol{y}_{2\ldots d})^\top - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) \big) \\
&\times y_1 \phi^+(y_1, \|\boldsymbol{\theta}^\star\|) \mathrm{d}y_1 \\
=\;& \int_0^{+\infty} \frac{e^{2y_1\theta_1^{\langle t\rangle}} - e^{-2y_1\theta_1^{\langle t\rangle}}}{e^{2y_1\theta_1^{\langle t\rangle}} + e^{-2y_1\theta_1^{\langle t\rangle}} + e^{2B(\boldsymbol{y}_{2\ldots d}) - 2\langle \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\rangle} + e^{-2B(\boldsymbol{y}_{2\ldots d}) + 2\langle \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\rangle}} \\
&\times y_1 \phi^+(y_1, \|\boldsymbol{\theta}^\star\|) \mathrm{d}y_1.
\end{aligned}$$

Hence, we have

$$\mathrm{sgn}\Big(\int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}) y_1 \phi^+(y_1, \|\boldsymbol{\theta}^\star\|) \mathrm{d}y_1 \Big) = \mathrm{sgn}(\theta_1^{\langle t\rangle}), \quad \forall \boldsymbol{y}_{2\ldots d}, B(\boldsymbol{y}_{2\ldots d}).$$

To prove Equation (A.13), according to Lemma 2.3, we have

$$\begin{aligned}
2\mathsf{p}^{\langle t+1\rangle} - 1 &= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star}(2\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle})-1)\\
&= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star}(\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle})+\mathsf{w}_d(-\boldsymbol{y}-\boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle})-1)\\
&= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star}\frac{e^{2\langle \boldsymbol{y},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+e^{-2\langle \boldsymbol{y},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+2e^{-2\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle}}{e^{2\langle \boldsymbol{y},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+e^{-2\langle \boldsymbol{y},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+e^{2\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+e^{-2\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle}}-1\\
&= \mathbb{E}_{\boldsymbol{y}\sim f_2^\star}\frac{e^{-2\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle}-e^{2\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle}}{e^{2\langle \boldsymbol{y},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+e^{-2\langle \boldsymbol{y},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+e^{2\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle}+e^{-2\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle}}.
\end{aligned}$$

Hence, we have

$$\operatorname{sgn}(1-2\mathsf{p}^{\langle t+1\rangle}) = \operatorname{sgn}(\langle \boldsymbol{a}^{\langle t\rangle},\boldsymbol{\theta}^{\langle t\rangle}\rangle).$$

This completes the proof of Lemma 2.4.

A.1.2.2 Proof of Lemma 2.5

Recall that the Population EM estimates for Model 4 satisfy the following update rules:

$$\boldsymbol{\theta}^{\langle t+1\rangle} = G_\theta(\boldsymbol{\theta}^{\langle t\rangle}, w_1^{\langle t\rangle}; \boldsymbol{\theta}^\star, w_1^\star) \quad \text{and} \quad w_1^{\langle t+1\rangle} = G_w(\boldsymbol{\theta}^{\langle t\rangle}, w_1^{\langle t\rangle}; \boldsymbol{\theta}^\star, w_1^\star).$$

Moreover, let $w_2 = 1 - w_1$ and note that the symmetric property of G_θ and G_w, i.e.,

$$\begin{aligned}
G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star) + G_\theta(-\boldsymbol{\theta}, w_2; \boldsymbol{\theta}^\star, w_1^\star) &= 0\\
G_w(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star) + G_w(-\boldsymbol{\theta}, w_2; \boldsymbol{\theta}^\star, w_1^\star) &= 1.
\end{aligned} \tag{A.16}$$

Hence, we just need to show

- For all $\langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star\rangle > 0, w_1 \in [0.5, 1)$, we have

$$\langle G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star), \boldsymbol{\theta}^\star\rangle > 0 \quad \text{and} \quad G_w(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star) > 0.5. \tag{A.17}$$

- For all $\langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star\rangle < 0, w_1 \in (0, 0.5]$, we have

$$\langle G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star), \boldsymbol{\theta}^\star\rangle < 0 \quad \text{and} \quad G_w(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star) < 0.5. \tag{A.18}$$

and then by a simple induction argument, it is straightforward to show Lemma 2.5 holds.

We first prove Equation (A.17) and then the claim of Equation (A.18) immediately follows due to Equation (A.16). Since for any orthogonal matrices $\boldsymbol{V}$, we have

$$\begin{aligned}\langle G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star), \boldsymbol{\theta}^\star\rangle &= \langle G_\theta(\boldsymbol{V}\boldsymbol{\theta}, w_1; \boldsymbol{V}\boldsymbol{\theta}^\star, w_1^\star), \boldsymbol{V}\boldsymbol{\theta}^\star\rangle \\ G_w(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star) &= G_w(\boldsymbol{V}\boldsymbol{\theta}, w_1; \boldsymbol{V}\boldsymbol{\theta}^\star, w_1^\star)\end{aligned}$$

Hence, the claim of Lemma 2.5 is invariant to rotation of the coordinates. Hence, without loss of generality, we assume that $\boldsymbol{\theta} = (\|\boldsymbol{\theta}\|, 0, 0, \ldots, 0)^\top$ and $\boldsymbol{\theta}^\star = (\theta_\parallel^\star, \theta_\perp^\star, 0, \ldots, 0)^\top$ with $\theta_\parallel^\star > 0$. To prove Equation (A.17), let us first show $G_w(\boldsymbol{\theta}, w; \boldsymbol{\theta}^\star, w_1^\star) > 0.5$. It is straightforward to show that

$$\begin{aligned}G_w(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star) &= \int \frac{w_1 e^{y\|\theta\|}}{w_1 e^{y\|\theta\|} + w_2 e^{-y\|\theta\|}} \left(w_1^\star \phi(y - \theta_\parallel^\star) + w_2^\star \phi(y + \theta_\parallel^\star)\right) \mathrm{d}y \\ &=: g_w(\|\boldsymbol{\theta}\|, w_1; \theta_\parallel^\star, w_1^\star).\end{aligned}$$

Hence, we just need to show that

$$g_w(\theta, w_1; \theta^\star, w_1^\star) > 0.5, \quad \forall w_1 \in [0.5, 1), w_1^\star \in (0.5, 1), \theta > 0, \theta^\star > 0. \tag{A.19}$$

Note that

$$\frac{\partial g_w(\theta, w_1; \theta^\star, w_1^\star)}{\partial w_1} = \int \frac{1}{\left(w_1 e^{y\theta} + w_2 e^{-y\theta}\right)^2} \left(w_1^\star \phi(y - \theta^\star) + w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y > 0.$$

Hence, we just need to show $g_w(\theta, 0.5; \theta^\star, w_1^\star) > 0.5$. Note that

$$\begin{aligned}g_w(\theta, 0.5; \theta^\star, w_1^\star) - 0.5 &= \int \frac{e^{y\theta}}{e^{y\theta} + e^{-y\theta}} \left(w_1^\star \phi(y - \theta^\star) + w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y - 0.5 \\ &= \int \frac{e^{y\theta} - e^{-y\theta}}{2(e^{y\theta} + e^{-y\theta})} \left(w_1^\star \phi(y - \theta^\star) + w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y \\ &= \int_{y \geq 0} \phi(y) e^{-\frac{(\theta^\star)^2}{2}} \cdot \left(\frac{(2w_1^\star - 1)\left(\cosh_y(\theta^\star + \theta) - \cosh_y(\theta^\star - \theta)\right)}{2\cosh_y(\theta)}\right) \mathrm{d}y \\ &> 0,\end{aligned}$$

where $\cosh_y(x) = \frac{1}{2}(e^{yx} + e^{-yx})$. Hence, Equation (A.19) holds.

Now we just need to show $\langle G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star), \boldsymbol{\theta}^\star\rangle > 0$. It is straightforward to show that all components of $G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star)$ are 0 except for the first two components

denoted as $\tilde{\theta}_1$ and $\tilde{\theta}_2$. For the second component $\tilde{\theta}_2$, we have

$$\begin{aligned}\tilde{\theta}_2 &= \theta_\perp^\star \int \frac{w_1 e^{y\|\theta\|} - w_2 e^{-y\|\theta\|}}{w_1 e^{y\|\theta\|} + w_2 e^{-y\|\theta\|}} \left(w_1^\star \phi(y - \theta_\|^\star) - w_2^\star \phi(y + \theta_\|^\star)\right) \mathrm{d}y \\ &= \theta_\perp^\star \cdot s(\|\theta\|, w_1; \theta_\|^\star, w_1^\star), \end{aligned} \tag{A.20}$$

where function $S : \mathbb{R}^4 \to \mathbb{R}$ is defined by

$$s(\theta, w_1; \theta^\star, w_1^\star) \triangleq \int \frac{w_1 e^{y\theta} - w_2 e^{-y\theta}}{w_1 e^{y\theta} + w_2 e^{-y\theta}} \left(w_1^\star \phi(y - \theta^\star) - w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y. \tag{A.21}$$

For the first component $\tilde{\theta}_1$, we have

$$\begin{aligned}\tilde{\theta}_1 &= \theta_\|^\star \int \frac{w_1 e^{y\|\theta\|} - w_2 e^{-y\|\theta\|}}{w_1 e^{y\|\theta\|} + w_2 e^{-y\|\theta\|}} \left(w_1^\star \phi(y - \theta_\|^\star) - w_2^\star \phi(y + \theta_\|^\star)\right) \mathrm{d}y \\ &\quad + \int \frac{w_1 e^{y\|\theta\|} - w_2 e^{-y\|\theta\|}}{w_1 e^{y\|\theta\|} + w_2 e^{-y\|\theta\|}} \left(w_1^\star (y - \theta_\|^\star)\phi(y - \theta_\|^\star) + w_2^\star (y + \theta_\|^\star)\phi(y + \theta_\|^\star)\right) \mathrm{d}y \\ &\overset{\text{(i)}}{=} \theta_\|^\star \cdot s(\|\theta\|, w_1; \theta_\|^\star, w_1^\star) \\ &\quad + \|\theta\| \int \frac{4 w_1 w_2}{\left(w_1 e^{y\|\theta\|} + w_2 e^{-y\|\theta\|}\right)^2} \left(w_1^\star \phi(y - \theta_\|^\star) + w_2^\star \phi(y + \theta_\|^\star)\right) \mathrm{d}y \\ &> \theta_\|^\star \cdot s(\|\theta\|, w_1; \theta_\|^\star, w_1^\star), \end{aligned} \tag{A.22}$$

where equation (i) holds due to partial integration. Hence, by Equation (A.20) and Equation (A.22) and $\theta_\|^\star > 0$, we have

$$\langle G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star), \boldsymbol{\theta}^\star \rangle > \|\theta^\star\|^2 \cdot s(\|\theta\|, w_1; \theta_\|^\star, w_1^\star).$$

Hence, we just need to show

$$s(\theta, w_1; \theta^\star, w_1^\star) > 0, \quad \forall \theta > 0, w_1 \in [0.5, 1], \theta^\star > 0, w_1^\star \in (0.5, 1). \tag{A.23}$$

For $w_1 = 0.5$, by Equation (A.20), we have

$$\begin{aligned} s(\theta, 0.5; \theta^\star, w_1^\star) &= \int \frac{e^{y\theta} - e^{-y\theta}}{e^{y\theta} + e^{-y\theta}} \left(w_1^\star \phi(y - \theta^\star) - w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y \\ &= \int_{y \geq 0} \frac{e^{y\theta} - e^{-y\theta}}{e^{y\theta} + e^{-y\theta}} \phi(y) e^{-\frac{(\theta^\star)^2}{2}} \left(e^{y\theta^\star} - e^{-y\theta^\star}\right) \mathrm{d}y > 0. \end{aligned} \tag{A.24}$$

For $w_1 \in (0.5, 1]$, by Equation (A.20) and taking derivative with respect to $w_1^\star$, we

have

$$\begin{aligned}\frac{\partial s(\theta, w_1; \theta^\star, w_1^\star)}{\partial w_1^\star} &= \int \frac{w_1 e^{y\theta} - w_2 e^{-y\theta}}{w_1 e^{y\theta} + w_2 e^{-y\theta}} \left(\phi(y-\theta^\star) + \phi(y+\theta^\star)\right) \mathrm{d}y \\ &= \int_{y\geq 0} \frac{2(w_1^2 - w_2^2)}{\left(w_1 e^{y\theta} + w_2 e^{-y\theta}\right)\left(w_1 e^{-y\theta} + w_2 e^{y\theta}\right)} \left(\phi(y-\theta^\star) + \phi(y+\theta^\star)\right) \mathrm{d}y \\ &> 0.\end{aligned}$$

Hence, we just need to show

$$s(\theta, w_1; \theta^\star, 0.5) \geq 0, \quad \forall \theta > 0, w_1 \in (0.5, 1], \theta^\star > 0. \tag{A.25}$$

Note that

$$\begin{aligned}2s(\theta, w_1; \theta^\star, 0.5) &= \int \frac{w_1 e^{y\theta} - w_2 e^{-y\theta}}{w_1 e^{y\theta} + w_2 e^{-y\theta}} \left(\phi(y-\theta^\star) - \phi(y+\theta^\star)\right) \mathrm{d}y \\ &= \int_{y\geq 0} \frac{w_1 w_2 (e^{2y\theta} - e^{-2y\theta})}{\left(w_1 e^{y\theta} + w_2 e^{-y\theta}\right)\left(w_1 e^{-y\theta} + w_2 e^{y\theta}\right)} \left(\phi(y-\theta^\star) - \phi(y+\theta^\star)\right) \mathrm{d}y \\ &\geq 0.\end{aligned}$$

Hence, we have Equation (A.25) holds. Combine with Equation (A.24), we have Equation (A.23) holds which completes the proof of this lemma.

A.1.3 Proofs omitted in Sections 2.4.3

A.1.3.1 Proof of Lemma 2.9

We remind the reader that due to Lemma 2.4, we assume that $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle > 0$ and $\langle \boldsymbol{a}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\langle 0 \rangle} \rangle \geq 0$ without loss of generality. Then it is straightforward to show that if $\|\boldsymbol{\theta}^{\langle 0 \rangle}\| = 0$, then

$$\boldsymbol{a}^{\langle t \rangle} = \boldsymbol{\theta}^{\langle t \rangle} = \mathbf{0}, \quad \forall t \geq 1.$$

Hence, we assume that $\|\boldsymbol{\theta}^{\langle 0 \rangle}\| > 0$. Since $\|\boldsymbol{\theta}^{\langle t \rangle}\|$ is rotation invariant, we apply the sequence of coordinate systems $\mathcal{A}$ we introduced in Section 2.4.3. Our goal is to prove the boundedness of $a_{\langle t \rangle,1}^{\langle t+1 \rangle}, a_{\langle t \rangle,2}^{\langle t+1 \rangle}, \theta_{\langle t \rangle,1}^{\langle t+1 \rangle}$, and $\theta_{\langle t \rangle,2}^{\langle t+1 \rangle}$. We start with bounding $a_{\langle t \rangle,2}^{\langle t+1 \rangle}$ and

$\theta^{\langle t+1\rangle}_{\langle t\rangle,2}$. According to Equation (2.42) we have

$$\begin{aligned} a^{\langle t+1\rangle}_{\langle t\rangle,2} &= \frac{\gamma^{\langle t+1\rangle}_{\langle t\rangle,2}(1-2\mathsf{p}^{\langle t+1\rangle})}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})} \leq \frac{\gamma^{\langle t+1\rangle}_{\langle t\rangle,2}}{2\mathsf{p}^{\langle t+1\rangle}} \overset{\text{(i)}}{=} \frac{\theta^{\star}_{\langle t\rangle,2}S(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})}{2\mathsf{p}^{\langle t+1\rangle}}, \\ \theta^{\langle t+1\rangle}_{\langle t\rangle,2} &= \frac{\gamma^{\langle t+1\rangle}_{\langle t\rangle,2}}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})} \overset{\text{(ii)}}{\leq} \frac{\gamma^{\langle t+1\rangle}_{\langle t\rangle,2}}{\mathsf{p}^{\langle t+1\rangle}} \overset{\text{(iii)}}{=} \frac{\theta^{\star}_{\langle t\rangle,2}S(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})}{\mathsf{p}^{\langle t+1\rangle}}, \end{aligned} \tag{A.26}$$

where Equalities (i) and (iii) are due to Equation (2.45). To obtain Inequality (ii) we used the following chain of arguments: According to Lemma 2.4, $\mathsf{p}^{\langle t\rangle} \leq 0.5$ for every t. Hence, $2(1-\mathsf{p}^{\langle t+1\rangle}) \geq 1$.

With exactly same calculation showed in Equation (2.55), we have

$$\mathsf{p}^{\langle t+1\rangle} = g_p(a^{\langle t\rangle}_i, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1}) > S(a^{\langle t\rangle}_i, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1}).$$

Together with Equation (A.26), we obtain

$$\begin{aligned} |a^{\langle t+1\rangle}_{\langle t\rangle,2}| &\leq \frac{|\theta^{\star}_{\langle t\rangle,2}|}{2} \leq \frac{\|\boldsymbol{\theta}^{\star}\|}{2}. \\ |\theta^{\langle t+1\rangle}_{\langle t\rangle,2}| &\leq \frac{|\theta^{\star}_{\langle t\rangle,2}|}{2(1-\mathsf{p}^{\langle t+1\rangle})} \leq \|\boldsymbol{\theta}^{\star}\|. \end{aligned} \tag{A.27}$$

Hence the only remaining step is to bound $a^{\langle t+1\rangle}_{\langle t\rangle,1}$ and $\theta^{\langle t+1\rangle}_{\langle t\rangle,1}$. To bound $a^{\langle t+1\rangle}_{\langle t\rangle,1}$ we consider two separate cases.

1. $a^{\langle t\rangle}_1 \geq \theta^{\star}_{\langle t\rangle,1} \geq 0, t \geq 0$: First note that according to Equation (2.42), we have

$$\begin{aligned} 0 \leq a^{\langle t+1\rangle}_{\langle t\rangle,1} &= \frac{\gamma^{\langle t+1\rangle}_{\langle t\rangle,1}(1-2\mathsf{p}^{\langle t+1\rangle})}{2\mathsf{p}^{\langle t+1\rangle}(1-\mathsf{p}^{\langle t+1\rangle})} \\ &= \frac{g_\gamma(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})(1-2g_p(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1}))}{2g_p(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})(1-g_p(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1}))} \\ &\leq \frac{g_\gamma(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})}{2g_p(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})}. \end{aligned} \tag{A.28}$$

Hence to bound $a^{\langle t+1\rangle}_{\langle t\rangle,1}$ we require a bound for

$$\frac{g_\gamma(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})}{2g_p(a^{\langle t\rangle}_1, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^{\star}_{\langle t\rangle,1})}.$$

Our next lemma provides such a bound.

Lemma A.1. If $a \geq x_{\theta^\star} \geq 0$ and $\theta \geq 0$, we have

$$\frac{g_\gamma(a, \theta, x_{\theta^\star})}{2g_p(a, \theta, x_{\theta^\star})} \leq \frac{a + \sqrt{\frac{2}{\pi}}}{2}.$$

We prove this lemma in the Appendix A.1.5.2. Combining Equation (A.28) and Lemma A.1 proves

$$(a_{\langle t\rangle,1}^{\langle t+1\rangle})^2 \leq \left(\frac{a_1^{\langle t\rangle} + \sqrt{\frac{2}{\pi}}}{2}\right)^2 \leq \frac{\|\boldsymbol{a}^{\langle t\rangle}\|^2 + \frac{2}{\pi}}{2}.$$

Therefore, combined with Equation (A.27), we have

$$\begin{aligned} \|\boldsymbol{a}^{\langle t+1\rangle}\|^2 &= \|\boldsymbol{a}_{\langle t\rangle}^{\langle t+1\rangle}\| = (a_{\langle t\rangle,1}^{\langle t+1\rangle})^2 + (a_{\langle t\rangle,2}^{\langle t+1\rangle})^2 \\ &\leq \frac{\|\boldsymbol{a}^{\langle t\rangle}\|^2 + \frac{2}{\pi}}{2} + \frac{\|\boldsymbol{\theta}^\star\|^2}{4} \\ &\leq \max\left(\|\boldsymbol{a}^{\langle t\rangle}\|^2, \frac{2}{\pi} + \frac{\|\boldsymbol{\theta}^\star\|^2}{2}\right). \end{aligned} \tag{A.29}$$

2. $a_1^{\langle t\rangle} < \theta^\star_{\langle t\rangle,1}$: Again according to Equation (2.42) we have

$$\begin{aligned} 0 \leq a_{\langle t\rangle,1}^{\langle t+1\rangle} &= \frac{\gamma_{\langle t\rangle,1}^{\langle t+1\rangle}(1 - 2\mathsf{p}^{\langle t+1\rangle})}{2\mathsf{p}^{\langle t+1\rangle}(1 - \mathsf{p}^{\langle t+1\rangle})} \\ &= \frac{g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})(1 - 2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))}{2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))}. \end{aligned}$$

We know from Lemma 2.4 that $\mathsf{p}^{\langle t+1\rangle} \leq 0.5$. In the range $0 < \mathsf{p}^{\langle t+1\rangle} \leq 0.5$,

$$\frac{(1 - 2\mathsf{p}^{\langle t+1\rangle})}{2\mathsf{p}^{\langle t+1\rangle}(1 - \mathsf{p}^{\langle t+1\rangle})}$$

is a positive decreasing function of $\mathsf{p}^{\langle t+1\rangle}$. Hence, if we a lower bound for g_p may lead to an upper bound for $a_{\langle t\rangle,1}^{\langle t+1\rangle}$. The following lemma provides such an upper bound.

Lemma A.2. If $a \geq x_{\theta^\star} \geq 0$ and $\theta \geq 0$, we have

$$g_p(a, \theta, x_{\theta^\star}) \geq \frac{1}{2}(1 - \Phi(a - x_{\theta^\star})) + \frac{1}{2}(1 - \Phi(a + x_{\theta^\star})).$$

If $0 \leq a < x_{\theta^\star}$ and $\theta \geq 0$, we have

$$g_p(a, \theta, x_{\theta^\star}) \geq \frac{1}{4}.$$

We prove this lemma in Appendix A.1.5.1.

By plugging $g_p(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}) = 0.25$ in Equation (A.30), we have

$$\begin{aligned} 0 \leq a_{\langle t \rangle,1}^{\langle t+1 \rangle} &= \frac{\gamma_{\langle t \rangle,1}^{\langle t+1 \rangle}(1 - 2\mathsf{p}^{\langle t+1 \rangle})}{2\mathsf{p}^{\langle t+1 \rangle}(1 - \mathsf{p}^{\langle t+1 \rangle})} \leq \frac{4}{3} g_\gamma(a_1^{\langle t \rangle}, \|\boldsymbol{\theta}^{\langle t \rangle}\|, \theta^\star_{\langle t \rangle,1}) \\ &\leq \frac{4}{3}\int |y| \phi^+(y, \theta^\star_{\langle t \rangle,1}) \mathrm{d}y \leq \frac{4}{3}\sqrt{\int y^2 \phi^+(y, \theta^\star_{\langle t \rangle,1}) \mathrm{d}y} \\ &= \frac{4}{3}\sqrt{1 + (\theta^\star_{\langle t \rangle,1})^2} \leq \frac{4}{3}\sqrt{1 + \|\boldsymbol{\theta}^\star\|^2}. \end{aligned} \tag{A.30}$$

Combining this with Equation (A.27), we obtain

$$\begin{aligned} \|\boldsymbol{a}^{\langle t+1 \rangle}\|^2 = \|\boldsymbol{a}_{\langle t \rangle}^{\langle t+1 \rangle}\|^2 &\leq \frac{16}{9}(1 + \|\boldsymbol{\theta}^\star\|^2) + \frac{\|\boldsymbol{\theta}^\star\|^2}{4} \\ &= \frac{16}{9} + \frac{73}{36}\|\boldsymbol{\theta}^\star\|^2. \end{aligned} \tag{A.31}$$

Therefore combining Equation (A.29) and Equation (A.31), we have

$$\|\boldsymbol{a}^{\langle t \rangle}\|^2 \leq \max\left(\|\boldsymbol{a}^{\langle 0 \rangle}\|^2, \frac{2}{\pi} + \frac{\|\boldsymbol{\theta}^\star\|^2}{2}, \frac{16}{9} + \frac{73}{36}\|\boldsymbol{\theta}^\star\|^2\right) = c_{U,1}^2 < \infty, \forall t \geq 0.$$

So far we have bounded $\{\|\boldsymbol{a}^{\langle t \rangle}\|\}_{t \geq 0}$ by $c_{U,1}$. Also, in Equation (A.27) we obtained an upper bound for $a_{\langle t \rangle,1}^{\langle t+1 \rangle}$. Our next step is to obtain an upper bound for $\theta_{\langle t \rangle,1}^{\langle t+1 \rangle}$. First note that

$$\theta_{\langle t \rangle,1}^{\langle t+1 \rangle} = \frac{\gamma_{\langle t \rangle,1}^{\langle t+1 \rangle}}{2\mathsf{p}^{\langle t+1 \rangle}(1 - \mathsf{p}^{\langle t+1 \rangle})}.$$

Hence, we have to find an upper bound for g_γ and a lower bound for g_p. Note that

$$\frac{\partial g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})}{\partial a_1^{\langle t\rangle}} = -\int \frac{2\|\boldsymbol{\theta}^{\langle t\rangle}\|}{(e^{y\|\boldsymbol{\theta}^{\langle t\rangle}\| - a_1^{\langle t\rangle}\|\boldsymbol{\theta}^{\langle t\rangle}\|} + e^{-y\|\boldsymbol{\theta}^{\langle t\rangle}\| + a_1^{\langle t\rangle}\|\boldsymbol{\theta}^{\langle t\rangle}\|})^2}\phi^+(y, \theta^\star_{\langle t\rangle,1})\mathrm{d}y \leq 0.$$

Therefore $\mathsf{p}^{\langle t+1\rangle} = g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})$ is a decreasing function of $a_1^{\langle t\rangle}$. Since $a_1^{\langle t\rangle} \leq \|\boldsymbol{a}^{\langle t\rangle}\| \leq c_{U,1}$, with Lemma A.2, we have $\forall t \geq 0$

$$\begin{aligned}
\mathsf{p}^{\langle t+1\rangle} &= g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) \geq g_p(c_{U,1}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) \\
&\geq \min\{\frac{1}{4}, \frac{1}{2}(1 - \Phi(c_{U,1} - \theta^\star_{\langle t\rangle,1})) + \frac{1}{2}(1 - \Phi(c_{U,1} + \theta^\star_{\langle t\rangle,1}))\} \\
&\geq \frac{1}{4}(1 - \Phi(c_{U,1} + \|\boldsymbol{\theta}^\star\|)) \triangleq c_{U,2} > 0.
\end{aligned}$$

Note that in Equation (A.30), we derived an upper bound for $g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})$. Therefore, we have

$$\begin{aligned}
0 \leq \theta^{\langle t+1\rangle}_{\langle t\rangle,1} &= \frac{\gamma^{\langle t+1\rangle}_{\langle t\rangle,1}}{2\mathsf{p}^{\langle t+1\rangle}(1 - \mathsf{p}^{\langle t+1\rangle})} \\
&\leq \frac{1}{2c_{U,2}(1 - c_{U,2})} g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) \leq \frac{1}{2c_{U,2}(1 - c_{U,2})}\sqrt{1 + \|\boldsymbol{\theta}^\star\|^2}
\end{aligned}$$

Thus with Equation (A.27), we have

$$\|\boldsymbol{\theta}^{\langle t\rangle}\|^2 \leq \max\{\|\boldsymbol{\theta}^{\langle 0\rangle}\|^2, \|\boldsymbol{\theta}^\star\|^2 + \frac{1}{4c_{U,2}^2(1 - c_{U,2})^2}(1 + \|\boldsymbol{\theta}^\star\|^2)\} = c_{U,3}^2 < \infty, \forall t \geq 0.$$

This completes the proof of Lemma 2.9.

A.1.3.2 Proof of Lemma 2.10

We remind the reader that due to Lemma 2.4, we assume that $\langle \boldsymbol{\theta}^{\langle 0\rangle}, \boldsymbol{\theta}^\star\rangle > 0$ and $\langle \boldsymbol{a}^{\langle 0\rangle}, \boldsymbol{\theta}^{\langle 0\rangle}\rangle \geq 0$ without loss of generality. Since $\|\boldsymbol{\theta}^{\langle t\rangle}\|$ is rotation invariant, we apply the sequence of coordinate systems $\mathcal{A}$ we introduced in Section 2.4.3. Note that we

have

$$\begin{aligned}\|\boldsymbol{\theta}^{\langle t+1\rangle}\| &= \|\boldsymbol{\theta}^{\langle t+1\rangle}_{\langle t\rangle}\| \geq \theta^{\langle t+1\rangle}_{\langle t\rangle,1} \\ &= \frac{g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})}{2g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})(1 - g_p(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}))} \\ &\geq 2g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}), \end{aligned} \tag{A.32}$$

where g_γ and g_p are defined in Equation (2.28) and Equation (2.27). Hence, the goal of the rest of the proof is to show that:

$$2g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) \geq \min\{\|\boldsymbol{\theta}^{\langle t\rangle}\|, c_l\}.$$

The main idea of this part is as follows. First note that

$$\frac{\partial g_\gamma(a, \theta, x_{\theta^\star})}{\partial \theta}\bigg|_{\theta=0} = \frac{1 + |x_{\theta^\star}|^2}{2}.$$

Hence, intuitively speaking we can argue that there exists a neighborhood of $\theta = 0$ on which the derivative is always larger than 0.5. Hence, when $\|\boldsymbol{\theta}^{\langle t\rangle}\|$ belongs to this neighborhood, $\|\boldsymbol{\theta}^{\langle t+1\rangle}\|$ is larger than $\|\boldsymbol{\theta}^{\langle t\rangle}\|$ and cannot go to zero. Next lemma justifies this claim.

Lemma A.3. For $\theta^\star_{\langle 0\rangle,1} > 0$ there exists a value δ_θ only depending on $c_{U,1}$, $\theta^\star_{\langle 0\rangle,1}$ and $\|\boldsymbol{\theta}^\star\|$ such that

$$\inf_{0\leq a\leq c_{U,1}, 0\leq\theta\leq\delta_\theta, \theta^\star_{\langle 0\rangle,1}\leq x_{\theta^\star}\leq\|\boldsymbol{\theta}^\star\|} \frac{\partial g_\gamma(a, \theta, x_{\theta^\star})}{\partial \theta} \geq 1/2.$$

We present the proof of this result in the Appendix A.1.5.4. We remind the reader that according to Lemma 2.9, $\|\boldsymbol{a}^{\langle t\rangle}\| \leq c_{U,1}$. Furthermore, since the angle $\beta^{\langle t\rangle}$ is a non increasing sequence according to Lemma 2.8, we have $\theta^\star_{\langle t\rangle,1} \geq \theta^\star_{\langle 0\rangle,1}$. Suppose that $\|\boldsymbol{\theta}^{\langle t\rangle}\| \leq \delta_\theta$. Then from Equation (A.32) we know that $\|\boldsymbol{\theta}^{\langle t+1\rangle}\| \geq 2g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1})$. Also from the mean value theorem we have:

$$|g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) - g_\gamma(a_1^{\langle t\rangle}, 0, \theta^\star_{\langle t\rangle,1})| = \frac{\partial g_\gamma(a_1^{\langle t\rangle}, \theta, \theta^\star_{\langle t\rangle,1})}{\partial \theta}\big|_{\theta=\xi} \cdot \|\boldsymbol{\theta}^{\langle t\rangle}\| \geq \frac{1}{2}\|\boldsymbol{\theta}^{\langle t\rangle}\|,$$

where $\xi \in [0, \|\boldsymbol{\theta}^{\langle t\rangle}\|]$ and to obtain the last inequality we used Lemma A.3 and the fact that $\|\boldsymbol{\theta}^{\langle t\rangle}\| \leq \delta_\theta$.

So far we have proved that if $\|\boldsymbol{\theta}^{\langle t\rangle}\| \leq \delta_\theta$, then $\|\boldsymbol{\theta}^{\langle t+1\rangle}\| \geq \|\boldsymbol{\theta}^{\langle t\rangle}\|$. But, we have not ruled out the possibility of the situation in which $\|\boldsymbol{\theta}^{\langle t\rangle}\| \geq \delta_\theta$, but $\|\boldsymbol{\theta}^{\langle t+1\rangle}\|$ is close to zero. That requires a simple continuity argument. Note that since g_γ is a continuous function of all its variables, its infimum over a compact set is achieved at certain point. Since, the value of $g_\gamma(x_1, \theta, x_{\theta^\star})$ is only zero when $\theta = 0$, we conclude that the infimum is not zero. Hence, we conclude that

$$c_l \triangleq \inf_{0\leq a\leq c_{U,1}, \delta_\theta \leq \theta \leq c_{U,3}, \theta^\star_{\langle 0\rangle,1} \leq x_{\theta^\star} \leq \|\boldsymbol{\theta}^\star\|} 2g_\gamma(a, \theta, x_{\theta^\star}) \;>\; 0.$$

Hence, we have if $\|\boldsymbol{\theta}^{\langle t\rangle}\| \geq \delta_\theta$, then

$$\|\boldsymbol{\theta}^{\langle t+1\rangle}\| \;\geq\; 2g_\gamma(a_1^{\langle t\rangle}, \|\boldsymbol{\theta}^{\langle t\rangle}\|, \theta^\star_{\langle t\rangle,1}) \;\geq\; c_l.$$

Therefore combining the result of $\|\boldsymbol{\theta}^{\langle t\rangle}\| \leq \delta_\theta$ and $\|\boldsymbol{\theta}^{\langle t\rangle}\| \geq \delta_\theta$, we know Lemma 2.10 holds.

A.1.3.3 Proof of Lemma 2.11

We consider three cases and deal with them separately: (i) $0 < a < x_{\theta^\star}$, (ii) $a \geq x_{\theta^\star}$, (iii) $a = 0$.

(i) $0 < a < x_{\theta^\star}$: Let $x_1 \triangleq x_{\theta^\star} + a > x_{\theta^\star} - a \triangleq x_2 > 0$. We first simplify the left hand side of the inequality in Equation (2.54). Our main goal in this section is to derive sharp upper bounds for $g_\gamma(a, \theta, x_{\theta^\star})$ and $1 - 2g_p(a, \theta, x_{\theta^\star})$. We start with $g_\gamma(a, \theta, x_{\theta^\star})$. Note that

$$\begin{aligned}
g_\gamma(a, \theta, x_{\theta^\star}) &= \int \mathsf{w}(y - a, \theta) y \phi^+(y, x_{\theta^\star}) \mathrm{d}y \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) + \int \left(\mathsf{w}(y - a, \theta) - \frac{1}{2} + \frac{1}{2}\right)(y - a)\phi^+(y, x_{\theta^\star})\mathrm{d}y \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) - \frac{1}{2}a + \int \left(\mathsf{w}(y - a, \theta) - \frac{1}{2}\right)(y - a)\phi^+(y, x_{\theta^\star})\mathrm{d}y \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) - \frac{1}{2}a \\
&\quad + \frac{1}{2}\int \left(\mathsf{w}(y - a, \theta) - \frac{1}{2}\right)(y - a)(\phi(y - x_{\theta^\star}) + \phi(y + x_{\theta^\star}))\mathrm{d}y \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) - \frac{1}{2}a + \frac{1}{4}\left(F(\theta, x_{\theta^\star} - a) + F(\theta, x_{\theta^\star} + a)\right) \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) - \frac{1}{2}a + \frac{1}{4}\left(F(\theta, x_1) + F(\theta, x_2)\right), \qquad \text{(A.33)}
\end{aligned}$$

where F is defined in Equation (2.31). Next, we find an upper bound for $\frac{1}{2}(F(\theta, x_1) + F(\theta, x_2))$. Note that $\forall \theta \geq 0, x_{\theta^\star} \geq 0$, we have

$$\begin{aligned}
F(\theta, x_{\theta^\star}) &= \int \frac{e^{y\theta} - e^{-y\theta}}{e^{y\theta} + e^{-y\theta}} y \frac{1}{\sqrt{2\pi}} e^{-(y - x_{\theta^\star})^2/2} \mathrm{d}y \\
&\leq \int_0^\infty y \frac{1}{\sqrt{2\pi}} e^{-(y - x_{\theta^\star})^2/2} \mathrm{d}y + \int_0^\infty y \frac{1}{\sqrt{2\pi}} e^{-(y + x_{\theta^\star})^2/2} \mathrm{d}y \\
&= x_{\theta^\star} \int_{-x_{\theta^\star}}^{x_{\theta^\star}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \mathrm{d}y + \int_{-x_{\theta^\star}}^\infty y \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \mathrm{d}y + \int_{x_{\theta^\star}}^\infty y \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \mathrm{d}y \\
&= x_{\theta^\star}(1 - 2\Phi(-x_{\theta^\star})) + 2\phi(x_{\theta^\star}) \triangleq l(x_{\theta^\star}). \qquad \text{(A.34)}
\end{aligned}$$

Therefore, if we replace $x_{\theta^\star}$ with x_1 and x_2 in Equation (A.34), we have

$$\frac{1}{2}\left(F(\theta, x_1) + F(\theta, x_2)\right) \leq \frac{1}{2}\left(l(x_1) + l(x_2)\right) \leq l(x_1), \qquad \text{(A.35)}$$

where the last inequality holds since $l(x)$ is an increasing function. This can be proved by taking the derivative of $l(x)$:

$$\frac{\mathrm{d}l(x_{\theta^\star})}{\mathrm{d}x_{\theta^\star}} = 1 - 2\Phi(-x_{\theta^\star}) + 2x_{\theta^\star}\phi(x_{\theta^\star}) - 2x_{\theta^\star}\phi(x_{\theta^\star}) = 1 - 2\Phi(-x_{\theta^\star}) \geq 0.$$

Combining Equation (A.33) and Equation (A.35) we obtain

$$g_\gamma(a, \theta, x_{\theta^\star}) \leq a \cdot g_p(a, \theta, x_{\theta^\star}) - \frac{1}{2}a + 2l(x_1). \qquad \text{(A.36)}$$

Now we obtain an upper bound for $1 - 2g_p(a, \theta, x_{\theta^\star})$. Note that,

$$\begin{aligned}
&1 - 2g_p(a, \theta, x_{\theta^\star}) \\
&= \int \left(\frac{1}{2} - \frac{e^{y\theta}}{e^{y\theta} + e^{-y\theta}}\right) \frac{1}{\sqrt{2\pi}} (e^{-(y + a - x_{\theta^\star})^2/2} + e^{-(y + a + x_{\theta^\star})^2/2}) \mathrm{d}y \\
&= \int \left(\frac{1}{2} - \frac{e^{y\theta}}{e^{y\theta} + e^{-y\theta}}\right) \frac{1}{\sqrt{2\pi}} (e^{-(y - x_2)^2/2} + e^{-(y + x_1)^2/2}) \mathrm{d}y \\
&= \int \frac{e^{-y\theta} - e^{y\theta}}{2(e^{y\theta} + e^{-y\theta})} \frac{1}{\sqrt{2\pi}} (e^{-(y - x_2)^2/2} + e^{-(y + x_1)^2/2}) \mathrm{d}y \\
&= \int \frac{e^{y\theta} - e^{-y\theta}}{2(e^{y\theta} + e^{-y\theta})} \frac{1}{\sqrt{2\pi}} (e^{-(y - x_1)^2/2} - e^{-(y - x_2)^2/2}) \mathrm{d}y \\
&= K(x_1, \theta) - K(x_2, \theta), \qquad \text{(A.37)}
\end{aligned}$$

where

$$K(x,\theta) \triangleq \int \frac{e^{y\theta} - e^{-y\theta}}{2(e^{y\theta} + e^{-y\theta})} \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2} \mathrm{d}y.$$

The following lemma proved in the Appendix A.1.5.5 summarizes some of the nice properties of this function, which will be used later in our proof.

Lemma A.4. $K(x,\theta)$ is a concave, strictly increasing function of x. Furthermore, $K(0,\theta) = 0$.

Given Equation (A.36) and Equation (A.37) we can now prove the claimed upper bound in Lemma 2.11. We have

$$\begin{aligned}
&\frac{g_\gamma(a,\theta,x_{\theta^\star})(1-2g_p(a,\theta,x_{\theta^\star}))}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))} \\
&= \frac{\left[a \cdot g_p(a,\theta,x_{\theta^\star}) - \frac{1}{2}a + \frac{1}{4}(F(\theta,x_1)+F(\theta,x_2))\right](1-2g_p(a,\theta,x_{\theta^\star}))}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))} \\
&= a + \frac{\frac{1}{2}(F(\theta,x_1)+F(\theta,x_2))(1-2g_p(a,\theta,x_{\theta^\star})) - a}{4g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))} \\
&= a + \frac{\frac{1}{2}(F(\theta,x_1)+F(\theta,x_2))(K(x_1,\theta)-K(x_2,\theta)) - \frac{x_1-x_2}{2}}{4g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))} \\
&\leq a + \frac{l(x_1)(K(x_1,\theta)-K(x_2,\theta)) - \frac{1}{2}(x_1-x_2)}{4g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))}.
\end{aligned} \tag{A.38}$$

It is straightforward to use the concavity of $K(x,\theta)$ in terms of x and prove that the function $\frac{K(x_1,\theta)-K(x_2,\theta)}{x_1-x_2}$ is a decreasing function of x_2. Hence, it is maximized at $x_2 = 0$. Since $K(0,\theta) = 0$, proved in Lemma A.4, we have

$$K(x_1,\theta) - K(x_2,\theta) \leq \frac{K(x_1,\theta)}{x_1}(x_1 - x_2). \tag{A.39}$$

Combining Equation (A.38) and Equation (A.39) implies:

$$\begin{aligned}\frac{g_\gamma(a,\theta,x_{\theta^\star})(1-2g_p(a,\theta,x_{\theta^\star}))}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))} &\leq a+\frac{\left(l(x_1)\frac{K(x_1,\theta)}{x_1}-\frac{1}{2}\right)(x_1-x_2)}{4g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))}\\ &= a+\frac{\left(l(x_1)\frac{K(x_1,\theta)}{x_1}-\frac{1}{2}\right)a}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))}\\ &= \left(1+\frac{l(x_1)\frac{K(x_1,\theta)}{x_1}-\frac{1}{2}}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))}\right)a.\end{aligned} \tag{A.40}$$

Our next step is to find an upper bound for $(l(x_1)\frac{K(x_1,\theta)}{x_1}-\frac{1}{2})$. Note that

$$\begin{aligned}K(x_1,\theta) &= \int\frac{e^{y\theta}-e^{-y\theta}}{2(e^{y\theta}+e^{-y\theta})}\frac{1}{\sqrt{2\pi}}e^{-(y-x_1)^2/2}\mathrm{d}y\\ &= \int_0^\infty\frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-(y-x_1)^2/2}\mathrm{d}y-\int_0^\infty\frac{e^{-y\theta}}{(e^{y\theta}+e^{-y\theta})}\frac{1}{\sqrt{2\pi}}e^{-(y-x_1)^2/2}\mathrm{d}y\\ &\quad-\int_0^\infty\frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-(y+x_1)^2/2}\mathrm{d}y+\int_0^\infty\frac{e^{-y\theta}}{(e^{y\theta}+e^{-y\theta})}\frac{1}{\sqrt{2\pi}}e^{-(y+x_1)^2/2}\mathrm{d}y\\ &\leq \int_0^{x_1}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\mathrm{d}y \;=\; \frac{1}{2}-\Phi(-x_1).\end{aligned}$$

Finally to obtain an upper bound for $\frac{1}{x}l(x)(\frac{1}{2}-\Phi(-x))$ we use the following lemma:

Lemma A.5. Define $l(x)\triangleq x(1-2\Phi(-x))+2\phi(x)$, then for all $x>0$, we have

$$\frac{l(x)(\frac{1}{2}-\Phi(-x))}{x}<\frac{1}{2}.$$

The proof of this lemma is presented in Appendix A.1.5.6. Using this lemma, we have

$$\frac{2l(x_1)(\frac{1}{2}-\Phi(-x_1))}{x_1}<1,\qquad \forall x_1=a+x_{\theta^\star}\in[x_{\theta^\star},2x_{\theta^\star}].$$

By continuity of the function $\frac{l(x)(\frac{1}{2}-\Phi(-x))}{x}$, we have

$$\bar{\kappa}_a(x_{\theta^\star})=\sup_{x_1\in[x_{\theta^\star},2x_{\theta^\star}]}\frac{2l(x_1)(\frac{1}{2}-\Phi(-x_1))}{x_1}\;<\;1.$$

It is straightforward to prove that $\bar{\kappa}_a(x_{\theta^\star})$ is a continuous function of $x_{\theta^\star} \in (0,\infty)$. Since $4g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star})) \leq 1$, we can bound Equation (A.40) in the following way:

$$\begin{aligned}
\frac{g_\gamma(a,\theta,x_{\theta^\star})(1-2g_p(a,\theta,x_{\theta^\star}))}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))} &\leq \left(1+\frac{l(x_1)\frac{\frac{1}{2}-\Phi(-x_1)}{x_1}-\frac{1}{2}}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))}\right)a \\
&\leq \left(1+\frac{\bar{\kappa}_a(x_{\theta^\star})-1}{4g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))}\right)a \\
&\leq \bar{\kappa}_a(x_{\theta^\star})a, \quad \forall 0<a<x_{\theta^\star}, \theta>0. \qquad \text{(A.41)}
\end{aligned}$$

This completes the proof of part (i).

(ii) $a \geq x_{\theta^\star}$: Note that

$$\begin{aligned}
1-2g_p(a,\theta,x_{\theta^\star}) &= \int (1-2\mathsf{w}(y-a,\theta))\,\phi^+(y,x_{\theta^\star})\mathrm{d}y \\
&= \int (1-\mathsf{w}(y-a,\theta)-\mathsf{w}(-y-a,\theta))\,\phi^+(y,x_{\theta^\star})\mathrm{d}y \\
&= \int \left(1-\frac{e^{2y\theta}+e^{-2y\theta}+2e^{-2a\theta}}{e^{2y\theta}+e^{-2y\theta}+e^{2a\theta}+e^{-2a\theta}}\right)\phi^+(y,x_{\theta^\star})\mathrm{d}y \\
&= \int \frac{e^{2a\theta}-e^{-2a\theta}}{e^{2y\theta}+e^{-2y\theta}+e^{2a\theta}+e^{-2a\theta}}\phi^+(y,x_{\theta^\star})\mathrm{d}y \\
&\geq 0. \qquad \text{(A.42)}
\end{aligned}$$

From Lemma A.1 with Equation (A.42), we have

$$\begin{aligned}
\frac{g_\gamma(a,\theta,x_{\theta^\star})(1-2g_p(a,\theta,x_{\theta^\star}))}{2g_p(a,\theta,x_{\theta^\star})(1-g_p(a,\theta,x_{\theta^\star}))} &\leq \frac{\left(a+\sqrt{\frac{2}{\pi}}\right)(1-2g_p(a,\theta,x_{\theta^\star}))}{2(1-g_p(a,\theta,x_{\theta^\star}))} \\
&= a+\frac{\sqrt{\frac{2}{\pi}}(1-2g_p(a,\theta,x_{\theta^\star}))-a}{2(1-g_p(a,\theta,x_{\theta^\star}))}. \qquad \text{(A.43)}
\end{aligned}$$

From Lemma A.2, we have

$$1-2g_p(a,\theta,x_{\theta^\star}) \leq \Phi(a+x_{\theta^\star})+\Phi(a-x_{\theta^\star})-1. \qquad \text{(A.44)}$$

Note that

$$\frac{\partial(\Phi(a+x_{\theta^\star})+\Phi(a-x_{\theta^\star})-1)}{\partial a} = \phi(a+x_{\theta^\star})+\phi(a-x_{\theta^\star}) \leq \sqrt{\frac{2}{\pi}}, \quad \forall a \in \mathbb{R},$$

and

$$\Phi(0 + x_{\theta^\star}) + \Phi(0 - x_{\theta^\star}) - 1 = 0.$$

Therefore, from Equation (A.44) and mean value theorem, we have

$$1 - 2g_p(a, \theta, x_{\theta^\star}) \leq \sqrt{\frac{2}{\pi}} a.$$

Together with Equation (A.43) and Equation (A.42), we have

$$\begin{aligned} \frac{g_\gamma(a, \theta, x_{\theta^\star})(1 - 2g_p(a, \theta, x_{\theta^\star}))}{2g_p(a, \theta, x_{\theta^\star})(1 - g_p(a, \theta, x_{\theta^\star}))} &\leq a + \frac{\sqrt{\frac{2}{\pi}}(1 - 2g_p(a, \theta, x_{\theta^\star})) - a}{2(1 - g_p(a, \theta, x_{\theta^\star}))} \\ &\leq a - \frac{1 - \frac{2}{\pi}}{2(1 - g_p(a, \theta, x_{\theta^\star}))} a \\ &\leq \left(\frac{1}{2} + \frac{1}{\pi}\right) a, \end{aligned} \tag{A.45}$$

where the last inequality holds due to the fact that $g_p(a, \theta, x_{\theta^\star}) \geq 0$.

(iii) $a = 0$: It is straightforward to prove that $g_p(0, \theta, x_{\theta^\star}) = \frac{1}{2}$. Hence,

$$\frac{g_\gamma(a, \theta, x_{\theta^\star})(1 - 2g_p(a, \theta, x_{\theta^\star}))}{2g_p(a, \theta, x_{\theta^\star})(1 - g_p(a, \theta, x_{\theta^\star}))} = 0.$$

Combining Case (i), (ii), (iii), we conclude that if we define

$$\kappa_a(x_{\theta^\star}) = \begin{cases} \max(\bar{\kappa}_a(x_{\theta^\star}), \frac{1}{2} + \frac{1}{\pi}), & x_{\theta^\star} > 0 \\ \frac{1}{2} + \frac{1}{\pi}, & x_{\theta^\star} = 0 \end{cases},$$

then the statement of Lemma 2.11 holds.

A.1.3.4 Proof of Lemma 2.13

According to the definition of function F in Equation (2.31), we have

$$\begin{aligned}\frac{\partial F(\theta, x_{\theta^\star}) - x_{\theta^\star}}{\partial \theta}\Big|_{\theta=x_{\theta^\star}} &= \int \frac{4y^2}{(e^{y\theta}+e^{-y\theta})^2}\frac{1}{\sqrt{2\pi}}e^{-(y-x_{\theta^\star})^2/2}\mathrm{d}y\Big|_{\theta=x_{\theta^\star}} \\ &= \int \frac{2y^2}{e^{yx_{\theta^\star}}+e^{-yx_{\theta^\star}}}\frac{1}{\sqrt{2\pi}}e^{-(y^2+x_{\theta^\star}^2)/2}\mathrm{d}y \\ &\leq e^{-\frac{x_{\theta^\star}^2}{2}}.\end{aligned}$$

We claim there exists $\delta > 0$ is a function of only $L_\theta, U_\theta, L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|$ such that $\forall|\theta - x_{\theta^\star}| \in [0,\delta], \theta \in [L_\theta, U_\theta], x_{\theta^\star} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$,

$$\frac{\partial F(\theta, x_{\theta^\star}) - x_{\theta^\star}}{\partial \theta} \leq \frac{1+e^{-\frac{L_{\theta^\star}^2}{2}}}{2}. \tag{A.46}$$

We prove it by contradiction. If not, for all $\delta > 0$, we have $\theta_\delta \in [L_\theta, U_\theta], \theta_\delta^\star \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|], |\theta_\delta - \theta_\delta^\star| \in [0,\delta]$ such that

$$\frac{\partial F(\theta, x_{\theta^\star}) - x_{\theta^\star}}{\partial \theta}\Big|_{\theta=\theta_\delta, x_{\theta^\star}=\theta_\delta^\star} > \frac{1+e^{-\frac{L_{\theta^\star}^2}{2}}}{2}.$$

For any sequence $\{\delta_i\}$ such that $\delta_i \to 0$, there exists subsequence δ_{i_j} such that $\{(\theta_{\delta_{i_j}}, \theta^\star_{\delta_{i_j}})\}$ converge to the limits $(\theta^\infty, \theta^\infty_\star)$. By compactness of the choice of $\theta, x_{\theta^\star}$, we have

$$\theta^\infty \in [L_\theta, U_\theta], \theta^\infty_\star \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|], |\theta^\infty - \theta^\infty_\star| \in [0, \lim_{j\to\infty}\delta_{i_j} = 0].$$

By continuity of $\frac{\partial F(\theta, x_{\theta^\star}) - x_{\theta^\star}}{\partial\theta}$, we have

$$\begin{aligned}\frac{1+e^{-\frac{L_{\theta^\star}^2}{2}}}{2} &\leq \lim_{j\to\infty}\frac{\partial F(\theta, x_{\theta^\star}) - x_{\theta^\star}}{\partial\theta}\Big|_{\theta=\theta_{\delta_{i_j}}, x_{\theta^\star}=\theta^\star_{\delta_{i_j}}} \\ &= \frac{\partial F(\theta, x_{\theta^\star}) - x_{\theta^\star}}{\partial\theta}\Big|_{\theta=x_{\theta^\star}=\theta^\infty_\star} \\ &= e^{-\frac{(\theta^\infty_\star)^2}{2}} < \frac{1+e^{-\frac{L_{\theta^\star}^2}{2}}}{2}.\end{aligned}$$

Contradiction! Hence we have Eq.(A.46) holds and for all $|\theta - x_{\theta^\star}| \in [0, \delta]$, $\theta \in [L_\theta, U_\theta]$, $x_{\theta^\star} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]$,

$$|F(\theta, x_{\theta^\star}) - x_{\theta^\star}| \;\leq\; |F(x_{\theta^\star}, x_{\theta^\star})| + \left| \frac{1 + e^{-\frac{L_{\theta^\star}^2}{2}}}{2} (\theta - x_{\theta^\star}) \right| \;=\; \frac{1 + e^{-\frac{L_{\theta^\star}^2}{2}}}{2} |\theta - x_{\theta^\star}|.$$

Note that from the definition of function F in Equation (2.31), we have

$$\begin{aligned} F(\theta, \theta^\star) &= \int \frac{e^{y\theta} - e^{-y\theta}}{e^{y\theta} + e^{-y\theta}} (y + \theta^\star) \frac{1}{\sqrt{2\pi}} e^{-(y-\theta^\star)^2/2} \mathrm{d}y \\ &= \int \frac{e^{y\theta} - e^{-y\theta}}{e^{y\theta} + e^{-y\theta}} (y + \theta^\star) \frac{1}{\sqrt{2\pi}} \left(\frac{1}{2} e^{-(y-\theta^\star)^2/2} + \frac{1}{2} e^{-(y-\theta^\star)^2/2} \right) \mathrm{d}y \\ &= H(\theta; \theta^\star, 0.5). \end{aligned} \tag{A.47}$$

With Equation (2.34), we have

$$|F(\theta, x_{\theta^\star}) - x_{\theta^\star}| < |\theta - x_{\theta^\star}|, \quad \forall |\theta - x_{\theta^\star}| \notin [0, \delta), \theta \in [L_\theta, U_\theta], x_{\theta^\star} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|].$$

Let

$$\kappa_b'' = \max \left(\frac{1 + e^{-\frac{L_{\theta^\star}^2}{2}}}{2}, \sup_{|\theta - x_{\theta^\star}| \notin [0,\delta), \theta \in [L_\theta, U_\theta], x_{\theta^\star} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|]} \frac{|F(\theta, x_{\theta^\star}) - x_{\theta^\star}|}{|\theta - x_{\theta^\star}|} \right),$$

by continuity of the function $\frac{|F(\theta, x_{\theta^\star}) - x_{\theta^\star}|}{|\theta - x_{\theta^\star}|}$, we have $\kappa_b'' \in (0, 1)$ is a function of only $L_\theta, U_\theta, L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|$ and

$$|F(\theta, x_{\theta^\star}) - x_{\theta^\star}| \leq \kappa_b'' |\theta - x_{\theta^\star}|, \quad \forall \theta \in [L_\theta, U_\theta], x_{\theta^\star} \in [L_{\theta^\star}, \|\boldsymbol{\theta}^\star\|].$$

This completes the proof of this lemma.

A.1.3.5 Proof of Lemma 2.7

In this lemma, we provide a new approach based on work in Tseng [2004] to prove convergence of $(\boldsymbol{a}^{\langle 0 \rangle}, \boldsymbol{\theta}^{\langle 0 \rangle})$ for the case $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle = 0$. From this approach, one can also show convergence of $(\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle})$ for the case $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle \neq 0$ as well. However, the strategy in Tseng [2004] neither can analyze the convergence speed nor where the estimates converges to directly. Further it is unclear whether it can be generalized to other GMM models. Hence, we do not adopt it for the proof of the case $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle \neq 0$.

The new approach uses the following strategy:

(i) We first characterize all the stationary points of Population EM. Let $(\boldsymbol{a}, \boldsymbol{\theta})$ denote the stationary points and we show that $\boldsymbol{a} = \mathbf{0}$ and $\boldsymbol{\theta} \in \{-\boldsymbol{\theta}^\star, \mathbf{0}, \boldsymbol{\theta}^\star\}$.

(ii) We then show that any accumulation point of $\{(\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle})\}$ is one of the stationary points. Let $(\boldsymbol{a}^\infty, \boldsymbol{\theta}^\infty)$ denote any accumulation point.

(iii) We show that if $\langle \boldsymbol{\theta}^{\langle 0 \rangle}, \boldsymbol{\theta}^\star \rangle = 0$, $\boldsymbol{\theta}^\infty$ can not converge to $-\boldsymbol{\theta}^\star$ or $\boldsymbol{\theta}^\star$. Hence, the algorithm has to converge to $\mathbf{0}$. Since $\boldsymbol{a}^\infty = \mathbf{0}$ for all stationary points, we have $\{\boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle}\}$ converges to $(\mathbf{0}, \mathbf{0})$.

A.1.3.6 Characterizing the Fixed Points of Population EM

First note that if we write the iterations of Population EM in terms of $\boldsymbol{a}^{\langle t \rangle}$ and $\boldsymbol{\theta}^{\langle t \rangle}$ we obtain

$$\begin{aligned} \boldsymbol{a}^{\langle t+1 \rangle} &= \frac{\boldsymbol{\gamma}^{\langle t+1 \rangle}(1 - 2\mathsf{p}^{\langle t+1 \rangle})}{2\mathsf{p}^{\langle t+1 \rangle}(1 - \mathsf{p}^{\langle t+1 \rangle})}, \\ \boldsymbol{\theta}^{\langle t+1 \rangle} &= \frac{\boldsymbol{\gamma}^{\langle t+1 \rangle}}{2\mathsf{p}^{\langle t+1 \rangle}(1 - \mathsf{p}^{\langle t+1 \rangle})}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\gamma}^{\langle t+1 \rangle} &= \int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle}) \boldsymbol{y} \phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}^\star) \mathrm{d}\boldsymbol{y}, \\ \mathsf{p}^{\langle t+1 \rangle} &= \int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle}) \phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}^\star) \mathrm{d}\boldsymbol{y}. \end{aligned}$$

If $(\boldsymbol{\gamma}^{\langle t \rangle}, \mathsf{p}^{\langle t \rangle}, \boldsymbol{a}^{\langle t \rangle}, \boldsymbol{\theta}^{\langle t \rangle})$ converges to $(\boldsymbol{\gamma}, \mathsf{p}, \boldsymbol{a}, \boldsymbol{\theta})$, then it is straightforward to show that

$$\boldsymbol{a} = \frac{\boldsymbol{\gamma}(1 - 2p)}{2p(1 - p)}, \tag{A.48}$$

$$\boldsymbol{\theta} = \frac{\boldsymbol{\gamma}}{2p(1 - p)}, \tag{A.49}$$

$$\boldsymbol{\gamma} = \int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}, \boldsymbol{\theta}) \boldsymbol{y} \phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}^\star) \mathrm{d}\boldsymbol{y}, \tag{A.50}$$

$$\mathsf{p} = \int \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}, \boldsymbol{\theta}) \phi_d^+(\boldsymbol{y}, \boldsymbol{\theta}^\star) \mathrm{d}\boldsymbol{y}. \tag{A.51}$$

Hence, the main step of the proof is to characterize the solutions of these four equations. We first consider the one-dimensional setting in which $Y \in \mathbb{R}$ and prove the

following two facts:

(i) The only feasible solution for a is zero.

(ii) We then set $a = 0$ and show that the only possible solutions for θ are $-\theta^\star, 0, \theta^\star$.

We should prove the above two by considering the following four different cases: (1) $a \geq 0, \theta \geq 0$, (2) $a \geq 0, \theta \leq 0$, (3) $a \leq 0, \theta \geq 0$, (4) $a \leq 0, \theta \leq 0$. Since the four cases are similar we focus on the first case only, i.e., $a \geq 0, \theta \geq 0$. To prove that the only possible solution of a is zero, note that Equation (A.48) can be written as

$$a = \frac{g_\gamma(a, \theta, \theta^\star)(1 - 2g_p(a, \theta, \theta^\star))}{2g_p(a, \theta, \theta^\star)(1 - g_p(a, \theta, \theta^\star))} \overset{(1)}{\leq} \kappa_a a. \tag{A.52}$$

where $\kappa_a < 1$. Note that Inequality (1) is a result of Lemma 2.11. Note that Equation (A.52) implies that a must be zero.

The only remaining step is to examine the solutions for θ. It is straightforward to prove that $g_p(0, \theta, \theta^\star) = \frac{1}{2}$. Hence, we can simplify Equation (A.49) to

$$\theta \;=\; 2g_\gamma(0, \theta, \theta^\star) \;=\; F(\theta, \theta^\star), \tag{A.53}$$

where the last equality is due to Equation (2.32). Hence, from Equation (2.34) and Equation (A.47), it proves our claim in the one dimensional setting.

To extend the proof to higher dimensions, we rotate the coordinates. Suppose that the fixed point is $\boldsymbol{a}, \boldsymbol{\theta}, \mathsf{p}, \boldsymbol{\gamma}$. Let $\tilde{\boldsymbol{M}}$ denote a rotation for which the following two hold: (i) $\tilde{\boldsymbol{\theta}} \triangleq \tilde{\boldsymbol{M}}\boldsymbol{\theta} = (\|\boldsymbol{\theta}\|_2, 0, 0, \ldots, 0)$ and (ii) $\tilde{\boldsymbol{\theta}}^\star \triangleq \tilde{\boldsymbol{M}}\boldsymbol{\theta}^\star = (\tilde{\theta}_1^\star, \tilde{\theta}_2^\star, 0, \ldots, 0)$. Define $\tilde{\boldsymbol{\gamma}} = \tilde{\boldsymbol{M}}\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{a}} = \tilde{\boldsymbol{M}}\boldsymbol{a}$. Lemma 2.3 shows that if we let $\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\gamma}}, \mathsf{p}$ denote the corresponding fixed point in the new coordinates, then they satisfy the same equations, i.e.,

$$\tilde{\boldsymbol{a}} \;=\; \frac{\tilde{\boldsymbol{\gamma}}(1 - 2\mathsf{p})}{2\mathsf{p}(1 - \mathsf{p})}, \tag{A.54}$$

$$\tilde{\boldsymbol{\theta}} \;=\; \frac{\tilde{\boldsymbol{\gamma}}}{2\mathsf{p}(1 - \mathsf{p})}, \tag{A.55}$$

$$\tilde{\boldsymbol{\gamma}} \;=\; \int \mathsf{w}_d(\boldsymbol{y} - \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\theta}})\boldsymbol{y}\phi_d^+(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}^\star)\mathrm{d}\boldsymbol{y}, \tag{A.56}$$

$$\mathsf{p} \;=\; \int \mathsf{w}_d(\boldsymbol{y} - \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\theta}})\phi_d^+(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}^\star)\mathrm{d}\boldsymbol{y}. \tag{A.57}$$

First, it is straightforward to employ Equation (A.54) and Equation (A.55) and con-

firm that $\forall i \geq 2$ we have

$$\begin{aligned}\tilde{\gamma}_i &= 2\tilde{\theta}_i \mathsf{p}(1-\mathsf{p}) = 0,\\ \tilde{a}_i &= \tilde{\theta}_i(1-2\mathsf{p}) = 0.\end{aligned}$$

Hence, with $\tilde{\boldsymbol{\theta}}_i^\star = 0, \forall i \geq 3$, we only need to consider the first two coordinates. Our goal is to prove the following two statements:

(i) If $\tilde{\theta}_2^\star \neq 0$, then $\tilde{\theta}_1 = 0$ and $\tilde{a}_1 = 0$. In other words, both $\boldsymbol{a}$ and $\boldsymbol{\theta}$ are zero.

(ii) If $\tilde{\theta}_2^\star = 0$, then the problem will be reduced to the one-dimensional problem that we have already discussed. Hence, it is straightforward to characterize the fixed points.

Here we focus on case (i), i.e., $\tilde{\theta}_2^\star \neq 0$. Since $\tilde{\gamma}_2 = 0$, we have

$$\begin{aligned}
0 = \tilde{\gamma}_2 &= \int \mathsf{w}_d(\boldsymbol{y}-\tilde{\boldsymbol{a}},\tilde{\boldsymbol{\theta}}) y_2 \phi_d^+(\boldsymbol{y},\tilde{\boldsymbol{\theta}}^\star)\mathrm{d}\boldsymbol{y}\\
&= \frac{1}{2}\int \mathsf{w}(y_1-\tilde{a}_1,\|\boldsymbol{\theta}\|)\phi(y_1-\tilde{\theta}_1^\star)\mathrm{d}y_1 \int y_2\phi(y_2-\tilde{\theta}_2^\star)\mathrm{d}y_2\\
&\quad + \frac{1}{2}\int \mathsf{w}(y_1-\tilde{a}_1,\|\boldsymbol{\theta}\|)\phi(y_1+\tilde{\theta}_1^\star)\mathrm{d}y_1 \int y_2\phi(y_2+\tilde{\theta}_2^\star)\mathrm{d}y_2\\
&= \tilde{\theta}_2^\star \int \mathsf{w}(y_1-\tilde{a}_1,\|\boldsymbol{\theta}\|)\phi^-(y_1,\tilde{\theta}_1^\star)\mathrm{d}y_1\\
&= \tilde{\theta}_2^\star \int_0^{+\infty} (\mathsf{w}(y_1-\tilde{a}_1,\|\boldsymbol{\theta}\|) - \mathsf{w}(-y_1-\tilde{a}_1,\|\boldsymbol{\theta}\|))\phi^-(y_1,\tilde{\theta}_1^\star)\mathrm{d}y_1\\
&= \tilde{\theta}_2^\star \int_0^{+\infty} \frac{e^{2y_1\|\boldsymbol{\theta}\|}-e^{-2y_1\|\boldsymbol{\theta}\|}}{e^{2y_1\|\boldsymbol{\theta}\|}+e^{-2y_1\|\boldsymbol{\theta}\|}+e^{2\|\tilde{\boldsymbol{a}}\|\|\boldsymbol{\theta}\|}+e^{-2\|\tilde{\boldsymbol{a}}\|\|\boldsymbol{\theta}\|}}\\
&\quad \times \frac{1}{2\sqrt{2\pi}}(e^{-(y_1-\tilde{\theta}_1^\star)^2/2}-e^{-(y_1+\tilde{\theta}_1^\star)^2/2})\mathrm{d}y_1.
\end{aligned}$$

It is straightforward to see that since $\tilde{\theta}_2^\star \neq 0$, then $\tilde{\theta}_1^\star = 0$. Hence, from Equation (A.54) and the definitions of g_γ and g_p functions given in Equation (2.27) and Equation (2.28) we have

$$\|\boldsymbol{a}\| = \|\tilde{\boldsymbol{a}}\| = |\tilde{a}_1| = |\tilde{\theta}_1||1-2p| = \frac{g_\gamma(\|\tilde{\boldsymbol{a}}\|,\|\boldsymbol{\theta}\|,0)(1-2g_p(\|\tilde{\boldsymbol{a}}\|,\|\boldsymbol{\theta}\|,0))}{2g_p(\|\tilde{\boldsymbol{a}}\|,\|\boldsymbol{\theta}\|,0)(1-g_p(\|\tilde{\boldsymbol{a}}\|,\|\boldsymbol{\theta}\|,0))} \leq \kappa_a\|\tilde{\boldsymbol{a}}\|,$$

where the last inequality is due to Lemma 2.11. We know that $\kappa_a < 1$. Therefore we

have $\boldsymbol{a} = \mathbf{0}$ and

$$\begin{aligned}
\|\boldsymbol{\theta}\| \;=\; \tilde{\theta}_1 &= \frac{\tilde{\gamma}_1}{2\mathsf{p}(1-\mathsf{p})} \\
&= 2\tilde{\gamma}_1 \\
&= 2\int \mathsf{w}(y-\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\theta}}) y_1 \phi_d^+(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}^\star) \mathrm{d}\boldsymbol{y} \\
&= 2\int \frac{e^{y_1\|\boldsymbol{\theta}\|}}{e^{y_1\|\boldsymbol{\theta}\|} + e^{-y_1\|\boldsymbol{\theta}\|}} y_1 \frac{1}{\sqrt{2\pi}} e^{-y_1^2/2} \mathrm{d}y \\
&= 0.
\end{aligned}$$

Thus the only solution is

$$(\boldsymbol{a}, \boldsymbol{\theta}) = (\mathbf{0}, \mathbf{0}).$$

A.1.3.7 Proof of Convergence for Population EM

We can break the proof into the following steps. Let $\{(\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\}_{t=1}^{\infty}$ denote all the estimates of the Population EM algorithm.

(i) We first prove the following lemma that analyzes any clustering point of the sequence $\{(\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\}$.

Lemma A.6. Any clustering point $(\boldsymbol{a}, \boldsymbol{\theta})$ of the estimates of the Population EM $\{(\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\}_t$ satisfy the following equations:

$$\begin{aligned}
\boldsymbol{a} &= \frac{\boldsymbol{\gamma}(1-2\mathsf{p})}{2\mathsf{p}(1-\mathsf{p})}, \\
\boldsymbol{\theta} &= \frac{\boldsymbol{\gamma}}{2\mathsf{p}(1-\mathsf{p})}, \\
\boldsymbol{\gamma} &= \mathbb{E}\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}, \boldsymbol{\theta})\boldsymbol{y}, \\
\mathsf{p} &= \mathbb{E}\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{a}, \boldsymbol{\theta}),
\end{aligned}$$

where $\boldsymbol{y} \sim 0.5N(-\boldsymbol{\theta}^\star, \boldsymbol{I}) + 0.5N(\boldsymbol{\theta}^\star, \boldsymbol{I})$.

We prove this lemma in Appendix A.1.5.7.

We have already proved that these fixed point equations only have the following solutions : $\boldsymbol{a} = \mathbf{0}$ and $\boldsymbol{\theta} \in \{-\boldsymbol{\theta}^\star, \mathbf{0}, \boldsymbol{\theta}^\star\}$. We proved in Lemma 2.4 that $\mathrm{sgn}(\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^\star\rangle) = \mathrm{sgn}(\langle \boldsymbol{\theta}^{\langle t+1\rangle}, \boldsymbol{\theta}^\star\rangle)$. Hence, we conclude that $\langle \boldsymbol{\theta}^{\langle t\rangle}, \boldsymbol{\theta}^\star\rangle = 0$ for every t. The only possible fixed point is hence $(\mathbf{0}, \mathbf{0})$. In summary, in the first step we prove that it only has one accumulation point, that is $(\mathbf{0}, \mathbf{0})$.

(ii) Next we prove that $\{\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\}_{t=1}^{\infty}$ is a convergent sequence. Suppose that the sequence does not converge to $(\mathbf{0}, \mathbf{0})$, then there exists an ϵ such that for every T, there exists a $t > T$ such that

$$\|(\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\|_2 > \epsilon.$$

We construct a subsequence of our sequence in the following way: Set $T = 1$ and pick $t_1 > T$ such that $\|(\boldsymbol{a}^{\langle t_1\rangle}, \boldsymbol{\theta}^{\langle t_1\rangle})\|_2 > \epsilon$. Now, set $T = t_1 + 1$, and pick $t_2 > T$ such that $\|(\boldsymbol{a}^{\langle t_2\rangle}, \boldsymbol{\theta}^{\langle t_2\rangle})\|_2 > \epsilon$. Continue the process until we construct a sequence $\{(\boldsymbol{a}^{\langle t_n\rangle}, \boldsymbol{\theta}^{\langle t_n\rangle})\}_{n=1}^{\infty}$. According to Lemma 2.9 $\{(\boldsymbol{a}^{\langle t_n\rangle}, \boldsymbol{\theta}^{\langle t_n\rangle})\}_{n=1}^{\infty}$ is in a compact set and has a convergent subsequence. But according to part (i) the converging subsequence of this sequence must converge to $(\mathbf{0}, \mathbf{0})$ which is in contradiction with the construction of the sequence $\{(\boldsymbol{a}^{\langle t_n\rangle}, \boldsymbol{\theta}^{\langle t_n\rangle})\}_{n=1}^{\infty}$. Hence $\{\boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}\}_{t=1}^{\infty}$ must be a convergent sequence and converges to $(\mathbf{0}, \mathbf{0})$.

A.1.4 Proofs omitted in Sections 2.4.4

A.1.4.1 Proof of Lemma 2.15

We study the shape of g_w by its first, second and third derivatives. Note that (with $w_2 = 1 - w_1$)

$$\frac{\partial g_w(\theta, w_1)}{\partial w_1} = \mathbb{E}_{y\sim f_4^\star}\left[\frac{1}{(w_1 e^{y\theta} + w_2 e^{-y\theta})^2}\right] > 0 \tag{A.58}$$

$$\frac{\partial^2 g_w(\theta, w_1)}{\partial w_1^2} = \mathbb{E}_{y\sim f_4^\star}\left[\frac{e^{-y\theta} - e^{y\theta}}{(w_1 e^{y\theta} + w_2 e^{-y\theta})^3}\right] \tag{A.59}$$

$$\frac{\partial^3 g_w(\theta, w_1)}{\partial w_1^3} = \mathbb{E}_{y\sim f_4^\star}\left[\frac{\left(e^{y\theta} - e^{-y\theta}\right)^2}{(w_1 e^{y\theta} + w_2 e^{-y\theta})^4}\right] > 0 \tag{A.60}$$

Hence, by Equation (A.60), we know the second derivative $\frac{\partial^2 g_w(\theta, w_1)}{\partial w_1^2}$ is a strictly increasing function of w_1 if $\theta \neq 0$. Hence, the second derivative can only change the sign at most once, the shape of g_w can only be one of the following three cases: (i) concave (the second derivative is always negative), (ii) concave-convex (the second derivative is negative, then positive) and (iii) convex (the second derivative is always positive). Note that by Lemma 2.5, we know $g_w(\theta, 0.5) > 0.5$ if $\theta > 0$. Moreover, it is easy to check that $g(\theta, 0) = 0$ and $g(\theta, 1) = 1$. Hence, we know for $\theta > 0$, the shape of g_w can only be either case (i) or case (ii). For case (i), it is clear that we have 1 is

the only stable fixed point and

$$g_w(\theta, w_1) > w_1 \quad \text{is equivalent to} \quad w_1 \in (0,1). \tag{A.61}$$

For case (ii), then depends on the value of the derivative at $w_1 = 1$ i.e., $\frac{\partial g_w(\theta,w_1)}{\partial w_1}|_{w_1=1}$, we have

- If $\partial g_w(\theta, w_1)/\partial w_1|_{w_1=1} \leq 1$, $w_1 = 1$ is the stable fixed point and Equation (A.61) holds.
- If $\partial g_w(\theta, w_1)/\partial w_1|_{w_1=1} < 1$, then $w_1 = 1$ is only a fixed point and there exists a stable fixed point in $(0,1)$ such that Equation (2.78) holds.

A.1.4.2 Proof of Lemma 2.16

We first analyze the condition that can determine the sign of $g(\theta, w_1) - w_1$. Note that (with $w_2 = 1 - w_1$)

$$\begin{aligned}
\frac{g(\theta, w_1) - w_1}{w_1} &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2+(\theta^\star)^2}{2}} \cdot \left(\frac{e^{y\theta}\left(w_1^\star e^{y\theta^\star} + w_2^\star e^{-y\theta^\star}\right)}{w_1 e^{y\theta} + w_2 e^{-y\theta}} - e^{y\theta^\star} \right) \mathrm{d}y \\
&= \int_{y\geq 0} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2+(\theta^\star)^2}{2}} \cdot \left(\frac{e^{y\theta}\left(w_1^\star e^{y\theta^\star} + w_2^\star e^{-y\theta^\star}\right)}{w_1 e^{y\theta} + w_2 e^{-y\theta}} - e^{y\theta^\star} \right. \\
&\quad \left. + \frac{e^{-y\theta}\left(w_1^\star e^{-y\theta^\star} + w_2^\star e^{y\theta^\star}\right)}{w_1 e^{-y\theta} + w_2 e^{y\theta}} - e^{-y\theta^\star} \right) \mathrm{d}y
\end{aligned}$$

Hence, to determine the sign of $g(\theta, w_1) - w_1 \gtrless 0$, we just need to show $\forall y \geq 0$

$$\left(\frac{e^{y\theta}\left(w_1^\star e^{y\theta^\star} + w_2^\star e^{-y\theta^\star}\right)}{w_1 e^{y\theta} + w_2 e^{-y\theta}} - e^{y\theta^\star} \right) + \left(\frac{e^{-y\theta}\left(w_1^\star e^{-y\theta^\star} + w_2^\star e^{y\theta^\star}\right)}{w_1 e^{-y\theta} + w_2 e^{y\theta}} - e^{-y\theta^\star} \right) \gtrless 0,$$

which is equivalent to

$$\begin{aligned}
0 \;\lessgtr\; & (2w_1 - 1)\cosh_y(\theta^\star) \\
& + (w_1^\star - w_1)\cosh_y(\theta^\star + 2\theta) + (1 - w_1 - w_1^\star)\cosh_y(\theta^\star - 2\theta)
\end{aligned} \tag{A.62}$$

where $\cosh_y(x) = (e^{yx} + e^{-yx})/2$. Let $\theta_\gamma = \gamma\theta^\star = \frac{2w_1^\star - 1}{2w_1 - 1}\theta^\star$. Let us first show that for $w_1 \in (0.5, 1]$

$$g_w(\theta_\gamma, w_1^\star) \;\gtrless\; w_1^\star, \quad \forall w_1 \gtrless w_1^\star. \tag{A.63}$$

By Equation (A.62), we just need to show

$$\cosh_y(\theta^\star) \;\geqslant\; \cosh_y(\theta^\star - 2\theta_\gamma), \quad \forall w_1 \gtrless w_1^\star,$$

which holds due to the monotonic of hyperbolic cosine function. Hence, we have proved Equation (A.63). Next, we want to show

$$g_w(\theta_\gamma, w_1) \;\gtrless\; w_1, \quad \forall w_1 \lesseqgtr w_1^\star. \tag{A.64}$$

By Equation (A.62), we just need to show that $\forall y > 0$,

$$\begin{aligned} 0 \;\lesseqgtr\; & (2w_1 - 1)\cosh_y(\theta^\star) + (w_1^\star - w_1)\cosh_y(\theta^\star + 2\theta_\gamma) \\ & + (1 - w_1 - w_1^\star)\cosh_y(\theta^\star - 2\theta_\gamma), \quad \forall w_1 \lesseqgtr w_1^\star. \end{aligned} \tag{A.65}$$

Note that, by Taylor expansion of $2\cosh_y(x) = \sum_{i=0}^{\infty} \frac{(xy)^{2i}}{(2i)!}$, we just need to show that given $\gamma = \frac{2w_1^\star - 1}{2w_1 - 1}$, we have

$$\begin{aligned} 0 \;<\; & (2w_1 - 1) + (w_1^\star - w_1)(1 + 2\gamma)^{2k} \\ & + (1 - w_1^\star - w_1)(2\gamma - 1)^{2k}, \qquad \forall w_1 \in (\frac{1}{2}, w_1^\star), k > 0, \end{aligned} \tag{A.66}$$

$$\begin{aligned} 0 \;<\; & (w_1 - w_1^\star)(1 + 2\gamma)^{2k} \\ & + (w_1^\star + w_1 - 1)(2\gamma - 1)^{2k} - (2w_1 - 1), \quad \forall w_1 \in (w_1^\star, 1], k > 1. \end{aligned} \tag{A.67}$$

For Equation (A.66), since $w_1 < w_1^\star$, we have $\gamma > 1$ and

$$\begin{aligned}
&(2w_1-1)+(w_1^\star-w_1)(1+2\gamma)^{2k}+(1-w_1^\star-w_1)(2\gamma-1)^{2k}\\
&= (w_1^\star-w_1)\left((1+2\gamma)^{2k}-(2\gamma-1)^{2k}\right)+(2w_1-1)\left(1-(2\gamma-1)^{2k}\right)\\
&= (w_1^\star-w_1)\cdot 2\left(\sum_{i=0}^{2k-1}(1+2\gamma)^i(2\gamma-1)^{2k-1-i}\right)\\
&\quad +(2w_1-1)\cdot(2\gamma-2)\left(\sum_{i=0}^{2k-1}(2\gamma-1)^i\right)\\
&= 2(w_1^\star-w_1)\left(\sum_{i=0}^{2k-1}\left((1+2\gamma)^i-2\right)(2\gamma-1)^{2k-1-i}\right)\\
&\geq 2(w_1^\star-w_1)\left(\sum_{i=0}^{1}\left((1+2\gamma)^i-2\right)(2\gamma-1)^{2k-1-i}\right)\\
&= 4(w_1^\star-w_1)(\gamma-1)(2\gamma-1)^{2k-2}>0.
\end{aligned}$$

For Equation (A.67), we have

$$\begin{aligned}
&(w_1-w_1^\star)(2\gamma+1)^{2k}+(w_1^\star+w_1-1)(2\gamma-1)^{2k}-(2w_1-1)\\
&= (w_1-w_1^\star)\left((2\gamma+1)^{2k}-(2\gamma-1)^{2k}\right)+(2w_1-1)\left((2\gamma-1)^{2k}-1\right)\\
&= (w_1-w_1^\star)\left((2\gamma+1)^2-(2\gamma-1)^2\right)\left(\sum_{i=0}^{k-1}(2\gamma+1)^{2i}(2\gamma-1)^{2k-2i-2}\right)\\
&\quad +(2w_1-1)\left((2\gamma-1)^2-1\right)\left(\sum_{i=0}^{k-1}(2\gamma-1)^{2i}\right)\\
&= 8(w_1-w_1^\star)\gamma\left(\sum_{i=0}^{k-1}\left((2\gamma+1)^{2i}-1\right)(2\gamma-1)^{2k-2i-2}\right)>0.
\end{aligned}$$

Hence, this completes the proof for Equation (A.64).

A.1.4.3 Proof of Lemma 2.17

We just need to bound $\|G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star)\|^2$. Note that by Equation (2.26) and Jensen's inequality, we have

$$\begin{aligned}\|G_\theta(\boldsymbol{\theta}, w_1; \boldsymbol{\theta}^\star, w_1^\star)\|^2 &\leq \mathbb{E}_{\boldsymbol{y}}\left[\left(\frac{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle} - w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}{w_1 e^{\langle \boldsymbol{y},\theta\rangle} + w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}\right)^2 \|\boldsymbol{y}\|^2\right] \\ &\leq \mathbb{E}_{\boldsymbol{y}}\|\boldsymbol{y}\|^2 = 1 + \|\boldsymbol{\theta}^\star\|^2.\end{aligned}$$

A.1.4.4 Proof of Lemma 2.18

To show Equation (2.82), we first define $\theta_\gamma = \gamma\theta^\star$, $\theta_b = b\theta^\star$, and

$$\begin{aligned}A &= \int y \frac{e^{y\theta_\gamma}}{w_1 e^{y\theta_\gamma} + (1-w_1)e^{-y\theta_\gamma}} \left(w_1^\star \phi(y-\theta^\star) + w_2^\star \phi(y+\theta^\star)\right) \mathrm{d}y \\ B &= \int y \frac{e^{-y\theta_\gamma}}{w_1 e^{y\theta_\gamma} + (1-w_1)e^{-y\theta_\gamma}} \left(w_1^\star \phi(y-\theta^\star) + w_2^\star \phi(y+\theta^\star)\right) \mathrm{d}y.\end{aligned}$$

Note that $\forall w_1$

$$(2w_1 - 1)\theta_\gamma \equiv w_1 A + w_2 B. \tag{A.68}$$

Hence, we have Equation (2.82) is equivalent to show that

$$w_1 A - w_2 B < \frac{w_1 A + w_2 B}{2w_1 - 1}, \quad \forall w_1 \in (0.5, w_1^\star),$$

which is equivalent to show

$$A + B > 0, \quad \forall w_1 \in (0.5, w_1^\star). \tag{A.69}$$

Note that

$$\begin{aligned}A + B &= \int \frac{1}{\sqrt{2\pi}} y(e^{y\theta_\gamma} + e^{-y\theta_\gamma}) e^{-\frac{y^2+(\theta^\star)^2}{2}} \frac{w_1^\star e^{y\theta^\star} + w_2^\star e^{-y\theta^\star}}{w_1 e^{y\theta_\gamma} + w_2 e^{-y\theta_\gamma}} \mathrm{d}y \\ &= \int_{y\geq 0} \frac{1}{\sqrt{2\pi}} y(e^{y\theta_\gamma} + e^{-y\theta_\gamma}) e^{-\frac{y^2+(\theta^\star)^2}{2}} \left(\frac{w_1^\star e^{y\theta^\star} + w_2^\star e^{-y\theta^\star}}{w_1 e^{y\theta_\gamma} + w_2 e^{-y\theta_\gamma}} - \frac{w_1^\star e^{-y\theta^\star} + w_2^\star e^{y\theta^\star}}{w_1 e^{-y\theta_\gamma} + w_2 e^{y\theta_\gamma}}\right) \mathrm{d}y \\ &= \int_{y\geq 0} \frac{1}{\sqrt{2\pi}} y(e^{y\theta_\gamma} + e^{-y\theta_\gamma}) e^{-\frac{y^2+(\theta^\star)^2}{2}} \\ &\quad \times \frac{(w_1^\star + w_1 - 1)\left(e^{y\theta^\star(1-\gamma)} - e^{-y\theta^\star(1-\gamma)}\right) + (w_1^\star - w_1)\left(e^{y\theta^\star(1+\gamma)} - e^{-y\theta^\star(1+\gamma)}\right)}{\left(w_1 e^{y\theta_\gamma} + w_2 e^{-y\theta_\gamma}\right)\left(w_1 e^{-y\theta_\gamma} + w_2 e^{y\theta_\gamma}\right)} \mathrm{d}y.\end{aligned}$$

Hence, we just need to show that for $\forall y > 0, w_1^\star, w_1 \in (\frac{1}{2}, 1)$,

$$\begin{aligned} 0 \;<\; & (w_1^\star + w_1 - 1)\left(e^{y\theta^\star(1-\gamma)} - e^{-y\theta^\star(1-\gamma)}\right) \\ & + (w_1^\star - w_1)\left(e^{y\theta^\star(1+\gamma)} - e^{-y\theta^\star(1+\gamma)}\right), \quad \forall w_1 \in (0.5, w_1^\star) \end{aligned}$$

By Taylor expansion of e^x, we just need to prove that for all $k \geq 0$, we have

$$(w_1^\star + w_1 - 1)(1-\gamma)^{2k+1} + (w_1^\star - w_1)(1+\gamma)^{2k+1} \;>\; 0, \quad \forall w_1 \in (0.5, w_1^\star)$$

By definition of γ, we just need to show

$$\begin{aligned} 0 \;<\; & (w_1^\star + w_1 - 1)2^{2k+1}(w_1 - w_1^\star)^{2k+1} \\ & + (w_1^\star - w_1)2^{2k+1}(w_1^\star + w_1 - 1)^{2k+1}, \quad \forall w_1 \in (0.5, w_1^\star) \\ \Leftrightarrow\; & w_1 + w_1^\star - 1 > w_1^\star - w_1, \quad \forall w_1 \in (0.5, w_1^\star), \end{aligned}$$

which obviously holds. To show Equation (2.83), we should analyze the condition for $g_\theta(\theta, w_1) - \theta > 0$. Note that

$$\begin{aligned} & g_\theta(\theta_b, w_1) - \theta_b \\ = \;& \int y \left(\frac{w_1 e^{y\theta_b} - w_2 e^{-y\theta_b}}{w_1 e^{y\theta_\gamma} + w_2 e^{-y\theta_\gamma}} - \frac{b}{w_1^\star - w_2^\star} \right) \left(w_1^\star \phi(y - \theta^\star) + w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y \\ = \;& \frac{1}{w_1^\star - w_2^\star} \int y \frac{w_1(2w_1^\star - 1 - b)e^{y\theta_b} - w_2(2w_1^\star - 1 + b)e^{-y\theta_b}}{w_1 e^{y\theta_\gamma} + w_2 e^{-y\theta_\gamma}} \\ & \times \left(w_1^\star \phi(y - \theta^\star) + w_2^\star \phi(y + \theta^\star)\right) \mathrm{d}y \\ = \;& \int_{y \geq 0} \frac{y}{\sqrt{2\pi}} e^{-\frac{y^2 + (\theta^\star)^2}{2}} \left(\frac{w_1 w_2 \left(2(1-b)\sinh_{y\theta^\star}(2b+1) + 2(1+b)\sinh_{y\theta^\star}(2b-1)\right)}{\left(w_1 e^{y\theta_\gamma} + w_2 e^{-y\theta_\gamma}\right)\left(w_1 e^{-y\theta_\gamma} + w_2 e^{y\theta_\gamma}\right)} \right. \\ & \left. + \frac{\left((2w_1 - 1)(2w_1^\star - 1) - (1 - 2w_1 w_2)\, b\right) \cdot 2\sinh_{y\theta^\star}(1)}{\left(w_1 e^{y\theta_\gamma} + w_2 e^{-y\theta_\gamma}\right)\left(w_1 e^{-y\theta_\gamma} + w_2 e^{y\theta_\gamma}\right)} \right) \mathrm{d}y, \end{aligned}$$

where $\sinh_{y\theta^\star}(x) = (e^{yx\theta^\star} - e^{-yx\theta^\star})/2$. Hence, we just need to show for all $y > 0$,

$$\begin{aligned} & w_1 w_2 \left((1-b)\sinh_{y\theta^\star}(2b+1) + (1+b)\sinh_{y\theta^\star}(2b-1)\right) \\ & \quad + \left((2w_1 - 1)(2w_1^\star - 1) - (1 - 2w_1 w_2)\, b\right)\sinh_{y\theta^\star}(1) \;> 0, \quad \forall b \in (0, \gamma], w_1 \in (w_1^\star, 1). \end{aligned}$$

By Taylor expansion of $\sinh_{y\theta^\star}(x)$, we just need to show for all $k \geq 0$, we have

$$\begin{aligned} & w_1 w_2 \left((1-b)(2b+1)^{2k+1} + (1+b)(2b-1)^{2k+1}\right) \\ & \quad + \left((2w_1-1)(2w_1^\star - 1) - (1-2w_1w_2)\, b\right) \ \geq 0, \quad \forall b \in (0,\gamma], w_1 \in (w_1^\star, 1). \end{aligned} \tag{A.70}$$

where inequality is strict for $k \geq 2$. It is straight forward to check Equation (A.70) holds for $k = 0$ due to $b \leq \gamma$. For $k \geq 1$, note that

$$\begin{aligned} & (1-b)(2b+1)^{2k+1} + (1+b)(2b-1)^{2k+1} \\ = \ & \left((2b+1)^{2k+1} + (2b-1)^{2k+1}\right) - b\left((2b+1)^{2k+1} - (2b-1)^{2k+1}\right) \\ = \ & 4b\sum_{i=0}^{2k}(-1)^i(2b+1)^{2k-i}(2b-1)^i - 2b\sum_{i=0}^{2k}(2b+1)^{2k-i}(2b-1)^i \\ = \ & 2b\left(\sum_{i=0}^{k-1}(2b+1)^{2k-2i-1}(2b-1)^{2i}(2b+1-3(2b-1)) + (2b-1)^{2k}\right) \\ = \ & 2b + 2b\left(\sum_{i=0}^{k-1}(2b+1)^{2k-2i-1}(2b-1)^{2i}(4-4b) + (2b-1)^{2k} - 1\right) \\ = \ & 2b + 2b(4-4b)\left(\sum_{i=0}^{k-1}(2b+1)^{2k-2i-1}(2b-1)^{2i} - \sum_{i=0}^{k-1}(2b-1)^{2i}b\right) \\ \geq \ & 2b + 2b(4-4b)\left(\sum_{i=0}^{k-1}(b+1)(2b-1)^{2i}\right) \\ \geq \ & 2b. \end{aligned}$$

where last two inequalities hold due to $b \leq \gamma < 1$ and last inequality is strict when $k \geq 2$. Hence, to show Equation (A.70), we just need to show

$$\begin{aligned} & 2bw_1w_2 + (2w_1-1)(2w_1^\star - 1) - (1-2w_1w_2)\, b \ \geq \ 0 \\ \Leftrightarrow \ & b \ \leq \ \gamma, \end{aligned}$$

which holds clearly. Hence, this completes the proof for this lemma.

A.1.5 Proof of Auxiliary Lemmas in Appendix A.1

A.1.5.1 Proof of Lemma A.2

According to the definition of $g_p(a, \theta, x_{\theta^\star})$ in Equation (2.27), we have

$$\begin{aligned} g_p(a,\theta,x_{\theta^\star}) &= \int \mathsf{w}(y-a,\theta)\phi^+(y,x_{\theta^\star})\mathrm{d}y \\ &= \int \mathsf{w}(y,\theta)\phi^+(y+a,x_{\theta^\star})\mathrm{d}y \\ &= \int_{y\geq 0} \phi^+(y+a,x_{\theta^\star})\mathrm{d}y + \int_{y\geq 0} \mathsf{w}(-y,\theta)(\phi^+(y-a,x_{\theta^\star}) - \phi^+(y+a,x_{\theta^\star}))\mathrm{d}y. \end{aligned} \tag{A.71}$$

where the last equality used the fact that $\mathsf{w}(y,\theta)+\mathsf{w}(-y,\theta) = 1$. If $a \geq x_{\theta^\star} \geq 0$, then

$$\begin{aligned} &2(\phi^+(y-a,x_{\theta^\star}) - \phi^+(y+a,x_{\theta^\star})) \\ &= \phi(y-a+x_{\theta^\star}) + \phi(y-a-x_{\theta^\star}) - \phi(y+a+x_{\theta^\star}) + \phi(y+a-x_{\theta^\star}) \\ &= \phi(y-a+x_{\theta^\star}) - \phi(y+a-x_{\theta^\star}) + \phi(y-a-x_{\theta^\star}) - \phi(y+a+x_{\theta^\star}) \\ &\geq 0, \qquad \forall y \geq 0. \end{aligned}$$

Hence with Equation (A.71), the above equation implies that if $a \geq x_{\theta^\star} \geq 0$, we have

$$\begin{aligned} g_p(a,\theta,x_{\theta^\star}) &\geq \int_{y\geq 0} \phi^+(y+a,x_{\theta^\star})\mathrm{d}y \\ &= \frac{1}{2}(1-\Phi(a-x_{\theta^\star})) + \frac{1}{2}(1-\Phi(a+x_{\theta^\star})). \end{aligned} \tag{A.72}$$

This completes the proof of the first part of Lemma A.2. Now we discuss the second part, i.e., the case $a < x_{\theta^\star}$. First note that g_p is a decreasing function of a, since $\theta \geq 0$ and

$$\frac{\partial g_p(a,\theta,x_{\theta^\star})}{\partial a} = -\int \frac{2\theta}{(e^{y\theta-a\theta}+e^{-y\theta+a\theta})^2}\phi^+(y,x_{\theta^\star})\mathrm{d}y \leq 0.$$

Therefore, we have

$$g_p(a,\theta,x_{\theta^\star}) \geq g_p(x_{\theta^\star},\theta,x_{\theta^\star}) \geq \frac{1}{4} + \frac{1}{2}(1-\Phi(2x_{\theta^\star}))$$

where the last inequality holds because $a = x_{\theta^\star}$ satisfies the condition of Equation (A.72). Hence, it immediately gives us that if $a < x_{\theta^\star}$, then

$$g_p(a, \theta, x_{\theta^\star}) \geq \frac{1}{4}.$$

This completes the proof.

A.1.5.2 Proof of Lemma A.1

We warn the reader that in this proof we use the proof of Lemma A.2, presented in the last section. According to the definition of g_γ in Equation (2.28), we have

$$\begin{aligned}
g_\gamma(a, \theta, x_{\theta^\star}) &= \int \mathsf{w}(y - a, \theta) y \phi^+(y, x_{\theta^\star}) \mathrm{d}y \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) + \int \mathsf{w}(y - a, \theta)(y - a) \phi^+(y, x_{\theta^\star}) \mathrm{d}y \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) + \int \mathsf{w}(y, \theta) y \phi^+(y + a, x_{\theta^\star}) \mathrm{d}y \\
&< a \cdot g_p(a, \theta, x_{\theta^\star}) + \int_{y \geq 0} y \phi^+(y + a, x_{\theta^\star}) \mathrm{d}y \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) + \frac{1}{2} \int_{y \geq 0} (y + a - x_{\theta^\star}) \phi(y + a - x_{\theta^\star}) \mathrm{d}y \\
&\quad + \frac{1}{2} \int_{y \geq 0} (y + a + x_{\theta^\star}) \phi(y + a + x_{\theta^\star}) \mathrm{d}y \\
&\quad - \frac{1}{2} \left((a - x_{\theta^\star}) \int_{y \geq 0} \phi(y + a - x_{\theta^\star}) \mathrm{d}y - (a + x_{\theta^\star}) \int_{y \geq 0} \phi(y + a + x_{\theta^\star}) \mathrm{d}y \right) \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) + \frac{1}{2} \left(\phi(x_{\theta^\star} - a) - (a - x_{\theta^\star})(1 - \Phi(a - x_{\theta^\star})) \right) \\
&\quad + \frac{1}{2} \left(\phi(x_{\theta^\star} + a) - (a + x_{\theta^\star})(1 - \Phi(a + x_{\theta^\star})) \right) \\
&= a \cdot g_p(a, \theta, x_{\theta^\star}) + \frac{1}{2} \left(W(a + x_{\theta^\star}) + W(a - x_{\theta^\star}) \right), \qquad \text{(A.73)}
\end{aligned}$$

where $W(x) = \phi(x) - x(1 - \Phi(x))$. Therefore we should find an upper bound for $W(x)$. Towards this goal we use the following lemma:

Lemma A.7. Let $\phi(x), \Phi(x)$ denote the pdf and CDF of standard Gaussian respectively. Then we have

$$\frac{\phi(x)}{1 - \Phi(x)} < x + \sqrt{\frac{2}{\pi}}, \quad \forall x > 0.$$

The proof of this Lemma is presented in Appendix A.1.5.3. Therefore from this lemma, we have

$$W(x) < \sqrt{\frac{2}{\pi}}(1 - \Phi(x)).$$

Hence we can upper bound Equation (A.73) by the following inequality:

$$g_\gamma(a, \theta, x_{\theta^\star}) \leq a \cdot g_p(a, \theta, x_{\theta^\star}) + \sqrt{\frac{2}{\pi}} \left(\frac{1}{2}(1 - \Phi(a + x_{\theta^\star})) + \frac{1}{2}(1 - \Phi(a - x_{\theta^\star})) \right).$$

From Lemma A.2, we have

$$g_p(a, \theta, x_{\theta^\star}) \geq \frac{1}{2}(1 - \Phi(a - x_{\theta^\star})) + \frac{1}{2}(1 - \Phi(a + x_{\theta^\star})), \quad \forall a \geq x_{\theta^\star}.$$

Therefore, if $a \geq x_{\theta^\star}$, then we have

$$\begin{aligned} g_\gamma(a, \theta, x_{\theta^\star}) &\leq a \cdot g_p(a, \theta, x_{\theta^\star}) + \sqrt{\frac{2}{\pi}} \left(\frac{1}{2}(1 - \Phi(a + x_{\theta^\star})) + \frac{1}{2}(1 - \Phi(a - x_{\theta^\star})) \right) \\ &\leq a \cdot g_p(a, \theta, x_{\theta^\star}) + \sqrt{\frac{2}{\pi}} g_p(a, \theta, x_{\theta^\star}), \end{aligned} \tag{A.74}$$

which completes the proof.

A.1.5.3 Proof of Lemma A.7

It is equivalent to show that

$$r(x) \triangleq (x + \sqrt{\frac{2}{\pi}})(1 - \Phi(x)) - \phi(x) > 0, \quad \forall x > 0.$$

Taking the first derivative of the left hand side, we have

$$\frac{\mathrm{d}r(x)}{\mathrm{d}x} = 1 - \Phi(x) - \phi(x)(x + \sqrt{\frac{2}{\pi}}) + x\phi(x) = 1 - \Phi(x) - \sqrt{\frac{2}{\pi}}\phi(x).$$

Taking the second derivative, we have

$$\frac{\mathrm{d}^2 r(x)}{\mathrm{d}x^2} = -\phi(x) + \sqrt{\frac{2}{\pi}} x\phi(x) = (\sqrt{\frac{2}{\pi}} x - 1)\phi(x).$$

Hence, we have $r''(x) < 0$ if $x < \sqrt{\pi/2}$ and $r''(x) > 0$ if $x > \sqrt{\pi/2}$. Therefore, $r'(x)$ is first strictly decreasing then strictly increasing function of x for $x \geq 0$. Since $r'(0) = 1/2 - 1/\pi > 0$, $r'(\sqrt{\pi/2}) = -0.04008391$ and

$$\lim_{x\to\infty} r'(x) \;=\; \lim_{x\to\infty} 1 - \Phi(x) - \sqrt{\frac{2}{\pi}}\phi(x) \;=\; 0,$$

we know there exists $x_0 \in (0, \sqrt{\pi/2})$ such that $r'(x) > 0$ if $x < x_0$ and $r'(x) < 0$ if $x > x_0$. Hence, $r(x)$ is first strictly increasing and then strictly decreasing function of $x \geq 0$. Since $r(0) = 0$ and

$$\begin{aligned}
|\lim_{x\to\infty} r(x)| &\leq \lim_{x\to\infty} \left(x + \sqrt{\frac{2}{\pi}}\right)(1 - \Phi(x)) + \phi(x) \\
&\leq 2\lim_{x\to\infty} \int_{y=x}^{+\infty} x\phi(y)\mathrm{d}y \\
&\leq 2\lim_{x\to\infty} \int_{y=x}^{+\infty} y\phi(y)\mathrm{d}y \\
&= 2\lim_{x\to\infty} \phi(x) = 0.
\end{aligned}$$

Hence, we have $r(x) > 0,\ \forall x > 0$. This completes the proof of this Lemma.

A.1.5.4 Proof of Lemma A.3

We first calculate the derivative $\frac{\partial g_\gamma(a,\theta,x_{\theta^\star})}{\partial\theta}$ at zero:

$$\begin{aligned}
\frac{\partial g_\gamma(a,\theta,x_{\theta^\star})}{\partial\theta}\Big|_{\theta=0} &= \int \frac{2(y-a)}{(e^{y\theta - a\theta} + e^{-y\theta + a\theta})^2} y\phi^+(y, x_{\theta^\star})\mathrm{d}y\Big|_{\theta=0} \\
&= \frac{1}{2}\int \frac{y^2 - ay}{2}(\phi(y - x_{\theta^\star}) + \phi(y + x_{\theta^\star}))\mathrm{d}y \\
&= \frac{1}{2}\left(1 + (x_{\theta^\star})^2\right).
\end{aligned}$$

This derivative is clearly larger than 0.5. Now we prove the main result by contradiction. Suppose that the claim of the lemma is not correct. Then, for any fixed $\{c_{U,1}, |\theta^\star_{\langle 0\rangle,1}|, \|\boldsymbol{\theta}^\star\|\}$, $\forall\delta > 0$, we have $a_\delta \in [0, c_{U,1}], \theta_\delta \in [0, \delta], \theta^\star_\delta \in [\theta^\star_{\langle 0\rangle,1}, \|\boldsymbol{\theta}^\star\|]$ such that

$$\frac{\partial g_\gamma(a,\theta,x_{\theta^\star})}{\partial\theta}\Big|_{a=a_\delta,\theta=\theta_\delta,x_{\theta^\star}=\theta^\star_\delta} < \frac{1}{2}.$$

Therefore, for any sequence $\{\delta_i\}$ such that $\delta_i \to 0$, we have

$$a_{\delta_i} \in [0, c_{U,1}], \quad \theta_{\delta_i} \in [0, \delta_i], \quad \theta^\star_{\delta_i} \in [\theta^\star_{\langle 0 \rangle,1}, \|\boldsymbol{\theta}^\star\|].$$

Since the sequence $\{a_{\delta_i}, \theta_{\delta_i}, \theta^\star_{\delta_i}\}_{i=1}^\infty$, belong to a compact set, there exists a subsequence δ_{i_j} such that $\{(a_{\delta_{i_j}}, \theta_{\delta_{i_j}}, \theta^\star_{\delta_{i_j}})\}$ converges to a limit $(a^\infty, \theta^\infty, \theta^\infty_\star)$ satisfying

$$a^\infty \in [0, c_{U,1}], \quad \theta^\infty \in [0, \lim_{j\to\infty} \delta_{i_j} = 0], \quad \theta^\infty_\star \in [\theta^\star_{\langle 0 \rangle,1}, \|\boldsymbol{\theta}^\star\|].$$

By continuity of $\frac{\partial g_\gamma(a,\theta,x_{\theta^\star})}{\partial \theta}$, we have

$$\begin{aligned}
\frac{1}{2} &\geq \lim_{j\to\infty} \frac{\partial g_\gamma(a,\theta,x_{\theta^\star})}{\partial \theta}\Big|_{a=a_{\delta_{i_j}},\theta=\theta_{\delta_{i_j}},x_{\theta^\star}=\theta^\star_{\delta_{i_j}}} \\
&= \frac{\partial g_\gamma(a,\theta,x_{\theta^\star})}{\partial \theta}\Big|_{a=a^\infty,\theta=0,x_{\theta^\star}=\theta^\infty_\star} \\
&= \frac{1}{2}(1+(\theta^\infty_\star)^2) \;>\; \frac{1}{2}.
\end{aligned}$$

This contradiction proves that Lemma A.3 is correct.

A.1.5.5 Proof of Lemma A.4

Firs note that if $x = 0$, then

$$K(0,\theta) = \int \frac{e^{y\theta} - e^{-y\theta}}{2(e^{y\theta} + e^{-y\theta})} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \mathrm{d}y,$$

which is the integral of an odd function and is hence equal to zero. To prove that the function is increasing and concave for $x \geq 0$, we calculate its derivatives. It is straightforward to see that

$$\begin{aligned}
\frac{\partial K(x,\theta)}{\partial x} &= \int \frac{e^{y\theta} - e^{-y\theta}}{2(e^{y\theta} + e^{-y\theta})} (y-x)\phi(y-x)\mathrm{d}y \\
&= -\int \frac{e^{y\theta} - e^{-y\theta}}{2(e^{y\theta} + e^{-y\theta})} \mathrm{d}\phi(y-x) \\
&= \int \phi(y-x) \frac{2\theta}{(e^{y\theta} + e^{-y\theta})^2} \mathrm{d}y > 0,
\end{aligned}$$

where the last equality is the result of integration by parts. Similarly, $\forall x \geq 0$

$$\begin{aligned}
\frac{\partial^2 K(x,\theta)}{\partial x^2} &= \int (y-x)\phi(y-x)\frac{2\theta}{(e^{y\theta}+e^{-y\theta})^2}\mathrm{d}y \\
&= -\int \frac{2\theta}{(e^{y\theta}+e^{-y\theta})^2}\mathrm{d}\phi(y-x) \\
&\overset{\text{(i)}}{=} -\int \phi(y-x)\frac{4\theta^2(e^{y\theta}-e^{-y\theta})}{(e^{y\theta}+e^{-y\theta})^3}\mathrm{d}y \\
&= \int_0^\infty (\phi(y+x)-\phi(y-x))\frac{4\theta^2(e^{y\theta}-e^{-y\theta})}{(e^{y\theta}+e^{-y\theta})^3}\mathrm{d}y < 0,
\end{aligned}$$

where equality (i) is an application of integration by parts.

A.1.5.6 Proof of Lemma A.5

Recall the definition of $l(x)$ in Equation (A.34) in Appendix A.1.3.3:

$$l(x) = x(1-2\Phi(-x)) + 2\phi(x).$$

Define

$$J(x) \;=\; \frac{1}{2}(x - l(x)(1-2\Phi(-x))) \;=\; 2\phi(x)\Phi(-x) + 2x\Phi(-x) - \phi(x) - 2x\Phi(-x)^2.$$

We would like to show that $J(x) \geq 0$. Hence, we analyze the shape of the function $J(x)$ by taking the derivatives, for all $x > 0$

$$\begin{aligned}
\frac{\mathrm{d}J(x)}{\mathrm{d}x} &= -2\phi(x)^2 + 2\Phi(-x) - x\phi(x) - 2\Phi(-x)^2 \\
\frac{\mathrm{d}^2 J(x)}{\mathrm{d}x^2} &= \phi(x)(4\phi(x)x - 3 + x^2 + 4\Phi(-x)) \\
\frac{\mathrm{d}J''(x)/\phi(x)}{\mathrm{d}x} &= 2x(1-2x\phi(x)) \geq 2x(1-\sqrt{\frac{2}{\pi}}) > 0.
\end{aligned}$$

Therefore $J''(x)/\phi(x)$ is an strictly increasing function of x. With $J''(0) < 0$ and $J''(10) > 0$, we have $J'(x)$ is first strictly decreasing then strictly increasing function of x. Since $J'(0) = 1/2 - 1/\pi > 0$ and $\lim_{x\to\infty} J'(x) = 0$, we know $J(x)$ achieves its minimum at either 0 or ∞. Since $J(0) = 0$ and $\lim_{x\to\infty} J(x) = 0$, we have

$$J(x) > 0, \quad \forall x > 0.$$

This completes the proof of this lemma.

A.1.5.7 Proof of Lemma A.6

Here is a summary of our strategy to prove this result. We first prove that $\|\boldsymbol{a}^{\langle t+1\rangle} - \boldsymbol{a}^{\langle t\rangle}\| \to 0$ and $\|\boldsymbol{\theta}^{\langle t+1\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| \to 0$ as $t \to \infty$. Then we use the following simple argument to prove that in fact the clustering points must satisfy the above fixed point equations. Suppose that $(\boldsymbol{a}, \boldsymbol{\theta})$ is an accumulation point. Then there is a subsequence $\{(\boldsymbol{a}^{\langle t_i\rangle}, \boldsymbol{\theta}^{\langle t_i\rangle})\}_{i=1}^{\infty}$ that converges to $(\boldsymbol{a}, \boldsymbol{\theta})$. Since we have $\|\boldsymbol{a}^{\langle t+1\rangle} - \boldsymbol{a}^{\langle t\rangle}\| \to 0$ and $\|\boldsymbol{\theta}^{\langle t+1\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| \to 0$, we can simply argue that $\{(\boldsymbol{a}^{\langle t_i+1\rangle}, \boldsymbol{\theta}^{\langle t_i+1\rangle})\}_{i=1}^{\infty}$ also converges to $(\boldsymbol{a}, \boldsymbol{\theta})$. We know that

$$\begin{aligned}
\boldsymbol{a}^{\langle t_i+1\rangle} &= \frac{\boldsymbol{\gamma}^{\langle t_i\rangle}(1 - 2\mathsf{p}^{\langle t_i\rangle})}{2\mathsf{p}^{\langle t_i\rangle}(1 - \mathsf{p}^{\langle t_i\rangle})}, \\
\boldsymbol{\theta}^{\langle t_i+1\rangle} &= \frac{\boldsymbol{\gamma}^{\langle t_i\rangle}}{2\mathsf{p}^{\langle t_i\rangle}(1 - \mathsf{p}^{\langle t_i\rangle})}, \\
\boldsymbol{\gamma}^{\langle t_i+1\rangle} &= \mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle})\boldsymbol{y} \\
\mathsf{p}^{\langle t_i+1\rangle} &= \mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{a}^{\langle t\rangle}, \boldsymbol{\theta}^{\langle t\rangle}).
\end{aligned}$$

By taking the limit $i \to \infty$ from both sides of the above equations we obtain the fixed point equations. Hence, the rest of the section is devoted to the proof of $\|\boldsymbol{a}^{\langle t+1\rangle} - \boldsymbol{a}^{\langle t\rangle}\| \to 0$ and $\|\boldsymbol{\theta}^{\langle t+1\rangle} - \boldsymbol{\theta}^{\langle t\rangle}\| \to 0$. The technique we us to prove this claim was first developed in Tseng [2004]. Since $\boldsymbol{a}^{\langle t\rangle} = (\boldsymbol{\mu}_1^{\langle t\rangle} + \boldsymbol{\mu}_2^{\langle t\rangle})/2$ and $\boldsymbol{\theta}^{\langle t\rangle} = (\boldsymbol{\mu}_2^{\langle t\rangle} - \boldsymbol{\mu}_1^{\langle t\rangle})/2$, we only need to prove that $\|\boldsymbol{\mu}_1^{\langle t+1\rangle} - \boldsymbol{\mu}_1^{\langle t\rangle}\| \to 0$ and $\|\boldsymbol{\mu}_2^{\langle t+1\rangle} - \boldsymbol{\mu}_2^{\langle t\rangle}\| \to 0$.

Define the following notion of distance between two parameter vectors:

$$D(\boldsymbol{\eta}, \boldsymbol{\nu}) = -\mathbb{E}f(z|\boldsymbol{y}; \boldsymbol{\nu}) \sum_z \ln\left(\frac{f(z|\boldsymbol{y}; \boldsymbol{\eta})}{f(z|\boldsymbol{y}; \boldsymbol{\nu})}\right),$$

where $f(\cdot)$ indicates corresponding pdf. Let $\boldsymbol{\mu}^{\langle t\rangle}$ be a shorthand for $(\boldsymbol{\mu}_1^{\langle t\rangle}, \boldsymbol{\mu}_2^{\langle t\rangle})$. As the first step of our proof we would like to show that $D(\boldsymbol{\mu}^{\langle t+1\rangle}, \boldsymbol{\mu}^{\langle t\rangle}) \to 0$. From

Equation (2.1), we have

$$\begin{aligned} Q_f(\boldsymbol{\eta}|\boldsymbol{\nu}) &= \mathbb{E}\sum_z f(z|\boldsymbol{y};\boldsymbol{\nu})\ln\left(f(z,\boldsymbol{y};\boldsymbol{\eta})\right) \\ &= \mathbb{E}\ln\left(f(\boldsymbol{y};\boldsymbol{\eta})\right) + \mathbb{E}\sum_z f(z|\boldsymbol{y};\boldsymbol{\nu})\ln\left(f(z|\boldsymbol{y};\boldsymbol{\eta})\right) \\ &= \mathbb{E}\ln\left(f(\boldsymbol{y};\boldsymbol{\eta})\right) - D(\boldsymbol{\eta},\boldsymbol{\nu}) + H(\boldsymbol{\nu},\boldsymbol{\nu}) \\ &= L(\boldsymbol{\eta}) - D(\boldsymbol{\eta},\boldsymbol{\nu}) + H(\boldsymbol{\nu},\boldsymbol{\nu}), \end{aligned}$$

where

$$\begin{aligned} L(\boldsymbol{\eta}) &\triangleq \mathbb{E}\ln\left(f(\boldsymbol{y}|\boldsymbol{\eta})\right) = \mathbb{E}\ln\left(\frac{1}{2}\phi_d(\boldsymbol{y}-\boldsymbol{\eta}_1)+\frac{1}{2}\phi_d(\boldsymbol{y}-\boldsymbol{\eta}_2)\right) \\ &\leq -\frac{d}{2}\ln 2\pi, \qquad \text{(A.75)} \\ H(\boldsymbol{\eta},\boldsymbol{\nu}) &\triangleq \mathbb{E}\sum_z f(z|\boldsymbol{y};\boldsymbol{\nu})\ln\left(f(z|\boldsymbol{y};\boldsymbol{\eta})\right). \qquad \text{(A.76)} \end{aligned}$$

Hence,

$$\boldsymbol{\mu}^{\langle t+1\rangle} = \text{argmax}_{\boldsymbol{\mu}'} Q_f(\boldsymbol{\mu}'|\boldsymbol{\mu}^{\langle t\rangle}) = \text{argmax}_{\boldsymbol{\mu}'}\{L(\boldsymbol{\mu}') - D(\boldsymbol{\mu}',\boldsymbol{\mu}^{\langle t\rangle})\}.$$

Note that every estimate of Population EM is obtained in a trade-off between maximizing the expected log-likelihood and minimizing the distance between the two consecutive estimates. First note that

$$L(\boldsymbol{\mu}^{\langle t+1\rangle}) - D(\boldsymbol{\mu}^{\langle t+1\rangle},\boldsymbol{\mu}^{\langle t\rangle}) \geq L(\boldsymbol{\mu}^{\langle t\rangle}) - D(\boldsymbol{\mu}^{\langle t\rangle},\boldsymbol{\mu}^{\langle t\rangle}) = L(\boldsymbol{\mu}^{\langle t\rangle}) \qquad \text{(A.77)}$$

Hence, $L(\boldsymbol{\mu}^{\langle t+1\rangle}) \geq L(\boldsymbol{\mu}^{\langle t\rangle}) + D(\boldsymbol{\mu}^{\langle t+1\rangle},\boldsymbol{\mu}^{\langle t\rangle})$. Therefore, $\{L(\boldsymbol{\mu}^{\langle t\rangle})\}$ is a non-decreasing sequence. Since according to Equation (A.75), $L(\boldsymbol{\mu})$ is upper bounded, thus $\{L(\boldsymbol{\mu}^{\langle t\rangle})\}_t$ converges. Also, according to Equation (A.77) we have

$$0 \leq D(\boldsymbol{\mu}^{\langle t+1\rangle},\boldsymbol{\mu}^{\langle t\rangle}) \leq L(\boldsymbol{\mu}^{\langle t+1\rangle}) - L(\boldsymbol{\mu}^{\langle t\rangle}) \to 0, \quad \text{as } t\to\infty.$$

This implies that

$$\{D(\boldsymbol{\mu}^{\langle t+1\rangle},\boldsymbol{\mu}^{\langle t\rangle})\} \to 0, \text{ as } t\to\infty.$$

Note that $D(\cdot,\cdot)$ is a measure of discrepancy between its two arguments. However, our goal is to show that the Euclidean distance between $\boldsymbol{\mu}^{\langle t\rangle}$ and $\boldsymbol{\mu}^{\langle t+1\rangle}$ goes to zero.

The rest of the proof is devoted to this claim. Since,

$$\begin{aligned}\boldsymbol{\mu}^{\langle t\rangle} &= \frac{\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t-1\rangle})\boldsymbol{y}}{\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t-1\rangle})},\\ \boldsymbol{\mu}^{\langle t+1\rangle} &= \frac{\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})\boldsymbol{y}}{\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})},\end{aligned}$$

in order to prove $\|\boldsymbol{\mu}^{\langle t+1\rangle}-\boldsymbol{\mu}^{\langle t\rangle}\|\to 0$, we should show that $\forall z\in\{1,0\}$

$$\|\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})\boldsymbol{y}-\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})\boldsymbol{y}\|\to 0,$$

and

$$|\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})-\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})|\to 0.$$

If we define $\psi(x)=-\ln x+x-1$, then

$$D(\boldsymbol{\eta},\boldsymbol{\nu})=\mathbb{E}\sum_z \psi\left(\frac{f(z|\boldsymbol{y};\boldsymbol{\eta})}{f(z|\boldsymbol{y};\boldsymbol{\nu})}\right)f(z|\boldsymbol{y};\boldsymbol{\nu}), \tag{A.78}$$

Since $\psi(x)>0$ for every value of $x>0$, the fact that $D(\boldsymbol{\mu}^{\langle t+1\rangle},\boldsymbol{\mu}^{\langle t\rangle})\to 0$ implies that $\forall z\in\{1,0\}$

$$\mathbb{E}\psi\left(\frac{f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})}{f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})}\right)f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle}) \;\to\; 0,\ \text{as } t\to\infty. \tag{A.79}$$

Hence, for all $z\in\{1,0\}$

$$\begin{aligned}&\mathbb{E}\psi\left(\frac{f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})}{f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})}\right)f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})\\ =\;&\mathbb{E}\left\{(f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})-f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle}))-\left(\ln f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})-\ln f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})\right)f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})\right\}\\ \overset{\text{(i)}}{=}\;&\mathbb{E}\frac{f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})}{2\xi^2}(f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})-f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle}))^2\\ \overset{\text{(ii)}}{\geq}\;&\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})(f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})-f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle}))^2\\ \geq\;&\mathbb{E}f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})(f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})-f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle}))^2\mathbb{I}(\|\boldsymbol{y}\|<M),\quad\forall t,M>0.\end{aligned}$$

where Equality (i) is the result of the Taylor expansion on $\ln X$ and ξ is a number between $f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle})$ and $f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})$ and Inequality (ii) holds for the fact that

$$f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t+1\rangle}),f(z|\boldsymbol{y};\boldsymbol{\mu}^{\langle t\rangle})\in(0,1),\quad\forall z\in\{1,0\}.$$

Hence, with Equation (A.79), we have for all $M > 0, z \in \{1, 0\}$,

$$\mathbb{E}f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle})(f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}))^2 \mathbb{I}(\|\boldsymbol{y}\| < M) \to 0, \text{ as } t \to \infty.$$

According to Lemma 2.9 $\{(\boldsymbol{a}^{\langle t_n \rangle}, \boldsymbol{\theta}^{\langle t_n \rangle})\}_{n=1}^{\infty}$ is in a compact set and hence so is $\{\boldsymbol{\mu}^{\langle t \rangle}\}$. Since $f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle})$ is a continuous function of $\boldsymbol{y}$ and $\boldsymbol{\mu}^{\langle t \rangle}$ with $f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}) > 0$ and compactness of $\{\boldsymbol{\mu}^{\langle t \rangle}\}$, there exists a constant c only depending on M such that

$$f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}) > c, \quad \forall \|\boldsymbol{y}\| < M, z \in \{1, 0\}, t > 0.$$

Therefore, for all $M > 0, z \in \{1, 0\}$, we have

$$\mathbb{E}(f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}))^2 \mathbb{I}(\|\boldsymbol{y}\| < M) \to 0, \text{ as } t \to \infty. \tag{A.80}$$

Also, for all $t \geq 0, z \in \{1, 0\}$, we have

$$\mathbb{E}(f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}))^2 \mathbb{I}(\|y\| \geq M) \leq \mathbb{E}\mathbb{I}(\|\boldsymbol{y}\| \geq M) \to 0, \text{ as } M \to \infty. \tag{A.81}$$

With Equation (A.80) and Equation (A.81), we have

$$\mathbb{E}(f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}))^2 \quad \to \quad 0, \text{ as } t \to \infty, \quad \forall z \in \{1, 0\}.$$

Therefore for all $z \in \{1, 0\}$, as $t \to \infty$, we have,

$$\begin{aligned} \|\mathbb{E}(f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}))\boldsymbol{y}\| &\leq \sqrt{\mathbb{E}(f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}))^2 \mathbb{E}\|\boldsymbol{y}\|^2} \\ &\to 0, \\ |\mathbb{E}f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle})| &\leq \sqrt{\mathbb{E}(f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle}) - f(z|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle}))^2} \\ &\to 0. \end{aligned}$$

Hence with compactness on sequence $\{\boldsymbol{\mu}^{\langle t \rangle}\}$, we have

$$\|\boldsymbol{\mu}_1^{\langle t+2 \rangle} - \boldsymbol{\mu}_1^{\langle t+1 \rangle}\| = \left\| \frac{\mathbb{E}f(z=0|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle})\boldsymbol{y}}{\mathbb{E}f(z=0|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle})} - \frac{\mathbb{E}f(z=0|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle})\boldsymbol{y}}{\mathbb{E}f(z=0|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle})} \right\| \to 0, \text{ as } t \to \infty,$$

and

$$\|\boldsymbol{\mu}_2^{\langle t+2 \rangle} - \boldsymbol{\mu}_2^{\langle t+1 \rangle}\| = \left\| \frac{\mathbb{E}f(z=1|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle})\boldsymbol{y}}{\mathbb{E}f(z=1|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t+1 \rangle})} - \frac{\mathbb{E}f(z=1|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle})\boldsymbol{y}}{\mathbb{E}f(z=1|\boldsymbol{y}; \boldsymbol{\mu}^{\langle t \rangle})} \right\| \to 0, \text{ as } t \to \infty.$$

This completes the proof of this lemma.

A.2 Proofs of Sample-based EM and Landscape results

A.2.1 Proofs omitted in Sections 2.5

A.2.1.1 Proof of Lemma 2.19

We first prove Equation (2.102) for $\|\hat{\boldsymbol{a}}^{\langle t\rangle}\|$. Clearly the result holds for $t = 0$. For all $t \geq 1$, using the condition Equation (2.100) on $\|\hat{\boldsymbol{a}}^{\langle t'\rangle}\|$ for all $t' \leq t$, we have

$$\begin{aligned}
\|\hat{\boldsymbol{a}}^{\langle t\rangle}\| &\leq \kappa_a \|\hat{\boldsymbol{a}}^{\langle t-1\rangle}\| + \epsilon_a \\
&\leq \kappa_a(\kappa_a \|\hat{\boldsymbol{a}}^{\langle t-2\rangle}\| + \epsilon_a) + \epsilon_a \\
&\leq (\kappa_a)^t \|\hat{\boldsymbol{a}}^{\langle 0\rangle}\| + \epsilon_a \sum_{i=0}^{t-1} (\kappa_a)^i \\
&\leq (\kappa_a)^t \|\hat{\boldsymbol{a}}^{\langle 0\rangle}\| + \frac{1}{1-\kappa_a}\epsilon_a.
\end{aligned}$$

Hence Equation (2.102) holds. Next, we prove Equation (2.103) for $\|\hat{\boldsymbol{\theta}}^{\langle t\rangle}\|$. Clearly the result holds for $t = 0$. For all $t \geq 1$, using the condition Equation (2.101) on $\|\hat{\boldsymbol{\theta}}^{\langle t'\rangle}\|$ for all $t' \leq t$, we have

$$\begin{aligned}
\|\hat{\boldsymbol{\theta}}^{\langle t\rangle} - \boldsymbol{\theta}^\star\| &\leq \kappa_\theta \|\hat{\boldsymbol{\theta}}^{\langle t-1\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_\theta \|\hat{\boldsymbol{a}}^{\langle t-1\rangle}\|} + \epsilon_\theta \\
&\leq \kappa_\theta(\kappa_\theta \|\hat{\boldsymbol{\theta}}^{\langle t-2\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_\theta \|\hat{\boldsymbol{a}}^{\langle t-2\rangle}\|} + \epsilon_\theta) + \epsilon_\theta \\
&\leq (\kappa_\theta)^t \|\hat{\boldsymbol{\theta}}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_\theta} \sum_{i=0}^{t-1} (\kappa_\theta)^{t-1-i} \sqrt{\|\hat{\boldsymbol{a}}^{\langle i\rangle}\|} + \epsilon_\theta \sum_{i=0}^{t-1} (\kappa_\theta)^i \\
&\leq (\kappa_\theta)^t \|\hat{\boldsymbol{\theta}}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_\theta} \sum_{i=0}^{t-1} (\kappa_\theta)^{t-1-i} \sqrt{\|\hat{\boldsymbol{a}}^{\langle i\rangle}\|} + \frac{1}{1-\kappa_\theta}\epsilon_\theta.
\end{aligned}$$

From Equation (2.102), we have $\forall t \geq 0$,

$$\sqrt{\|\hat{\boldsymbol{a}}^{\langle t\rangle}\|} \leq \sqrt{(\kappa_a)^t \|\hat{\boldsymbol{a}}^{\langle 0\rangle}\| + \frac{1}{1-\kappa_a}\epsilon_a} \leq (\kappa_a)^{\frac{t}{2}} \sqrt{\|\hat{\boldsymbol{a}}^{\langle 0\rangle}\|} + \sqrt{\frac{1}{1-\kappa_a}\epsilon_a}.$$

Hence we have

$$\begin{aligned}
\|\hat{\boldsymbol{\theta}}^{\langle t\rangle}-\boldsymbol{\theta}^{\star}\| &\leq (\kappa_\theta)^t\|\hat{\boldsymbol{\theta}}^{\langle 0\rangle}-\boldsymbol{\theta}^{\star}\|+\sqrt{c_\theta}\sum_{i=0}^{t-1}(\kappa_\theta)^{t-1-i}\sqrt{\|\hat{\boldsymbol{a}}^{\langle i\rangle}\|}+\frac{1}{1-\kappa_\theta}\epsilon_\theta\\
&\leq (\kappa_\theta)^t\|\hat{\boldsymbol{\theta}}^{\langle 0\rangle}-\boldsymbol{\theta}^{\star}\|+\sqrt{c_\theta}\sum_{i=0}^{t-1}(\kappa_\theta)^{t-1-i}((\sqrt{\kappa_a})^i\sqrt{\|\hat{\boldsymbol{a}}^{\langle 0\rangle}\|}+\sqrt{\frac{1}{1-\kappa_a}\epsilon_a})+\frac{1}{1-\kappa_\theta}\epsilon_\theta\\
&= (\kappa_\theta)^t\|\hat{\boldsymbol{\theta}}^{\langle 0\rangle}-\boldsymbol{\theta}^{\star}\|+\sqrt{c_\theta\|\hat{\boldsymbol{a}}^{\langle 0\rangle}\|}\sum_{i=0}^{t-1}(\kappa_\theta)^{t-1-i}(\sqrt{\kappa_a})^i\\
&\quad+\sqrt{\frac{c_\theta}{1-\kappa_a}\epsilon_a}\sum_{i=0}^{t-1}(\kappa_\theta)^i+\frac{1}{1-\kappa_\theta}\epsilon_\theta\\
&\leq (\kappa_\theta)^t\|\hat{\boldsymbol{\theta}}^{\langle 0\rangle}-\boldsymbol{\theta}^{\star}\|+t\sqrt{c_\theta\|\hat{\boldsymbol{a}}^{\langle 0\rangle}\|}(\max\{\sqrt{\kappa_a},\kappa_\theta\})^t+\frac{1}{1-\kappa_\theta}\sqrt{\frac{c_\theta}{1-\kappa_a}\epsilon_a}+\frac{1}{1-\kappa_\theta}\epsilon_\theta.
\end{aligned}$$

This completes the proof of this lemma.

A.2.1.2 Proof of Equation (2.106)-Equation (2.108)

Lemma A.8. Let $y_1,\cdots,y_n \overset{i.i.d.}{\sim} \frac{1}{2}N(\boldsymbol{\theta}^{\star},\boldsymbol{I}_d)+\frac{1}{2}N(-\boldsymbol{\theta}^{\star},\boldsymbol{I}_d)$. Then, we have

(1) $\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\|\leq 4(\|\boldsymbol{\theta}^{\star}\|+1)\sqrt{\frac{2d+\ln(1/\delta)}{n}}$, with probability at least $1-\delta$.

(2) $\sup_{\substack{\|\boldsymbol{x}_\theta\|\leq c,\\ \|\boldsymbol{x}_a\|\leq 1}}|\frac{1}{n}\sum_{i=1}^n \mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\mathbb{E}\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{x}_a,\boldsymbol{x}_\theta)|\leq 8c(\|\boldsymbol{\theta}^{\star}\|+2)\sqrt{\frac{d+2+\ln(1/\delta)}{n}}$, with probability at least $1-\delta$.

(3) $\sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1}\|\frac{1}{n}\sum_{i=1}^n(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\boldsymbol{y}_i-\mathbb{E}(\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\boldsymbol{y}\|\leq 36c(\|\boldsymbol{\theta}^{\star}\|+2)\sqrt{\frac{d+2+\ln(1/\delta)}{n}}$, with probability at least $1-\delta$.

Proof. We first prove the first claim (1). Note that $\boldsymbol{y}_i$ can be expressed by $\boldsymbol{y}_i=\zeta_i\boldsymbol{\theta}^{\star}+\boldsymbol{\omega}_i$, where ζ_i are i.i.d sequence of Rademacher variables and $\boldsymbol{\omega}_i$ are i.i.d $N(\boldsymbol{0},\boldsymbol{I}_d)$ Gaussian random variables. Therefore we have,

$$\begin{aligned}
\|\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i\|^2 &= \|\frac{1}{n}\sum_{i=1}^n\zeta_i\boldsymbol{\theta}^{\star}+\frac{1}{n}\sum_{i=1}^n\boldsymbol{\omega}_i\|^2\\
&= \frac{1}{n}\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\zeta_i\boldsymbol{\theta}^{\star}+\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\omega}_i\|^2.
\end{aligned}$$

Note that $\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\omega}_i\|^2\overset{dist.}{=}\boldsymbol{\nu}$, where $\boldsymbol{\nu}\sim\chi^2(d)$. Hence, using Cramér-Chernoff

inequality, we have probability at least $1 - \frac{\delta}{2}$ such that

$$\left| \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\omega}_i\|^2 - d \right| \leq \sqrt{8d \ln(2/\delta)} \leq d + 2\ln(2/\delta) \text{ for sufficiently large } n.$$

Moreover, for Rademacher variables ζ_i, using Hoeffding's inequality, we have with probability at least $1 - \frac{\delta}{2}$ such that

$$|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i| \leq \sqrt{2\ln(2/\delta)}$$

Therefore, we have probability at least $1 - \delta$ such that

$$\begin{aligned}
\|\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{y}_i\| &\leq \frac{1}{\sqrt{n}} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i \boldsymbol{\theta}^\star + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\omega}_i\| \\
&\leq \frac{1}{\sqrt{n}} \sqrt{2\|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i \boldsymbol{\theta}^\star\|^2 + 2\|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\omega}_i\|^2} \\
&\leq \frac{1}{\sqrt{n}} \sqrt{2(2\ln(2/\delta))\|\boldsymbol{\theta}^\star\|^2 + 2(2d + 2\ln 2/\delta)} \\
&= 2\sqrt{\frac{\ln(2/\delta)(\|\boldsymbol{\theta}^\star\|^2 + 1) + d}{n}} \\
&\leq 4(\|\boldsymbol{\theta}^\star\| + 1)\sqrt{\frac{2d + \ln(1/\delta)}{n}}.
\end{aligned}$$

For the second claim, define

$$Z_+ \triangleq \sup_{\|\boldsymbol{x}_\theta\| \leq c, \|\boldsymbol{x}_a\| \leq 1} \frac{1}{n} \sum_{i=1}^{n} \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{x}_a, \boldsymbol{x}_\theta) - \mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta).$$

Then we have $\forall \|\boldsymbol{x}_\theta\| \leq c, \|\boldsymbol{x}_a\| \leq 1$

$$\begin{aligned}
\mathbb{E}e^{\lambda Z_+} &\overset{(i)}{\leq} \mathbb{E}_{\boldsymbol{y},\boldsymbol{y}'} e^{\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} \frac{1}{n}\sum_{i=1}^n (\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\mathsf{w}_d(\boldsymbol{y}_i'-\boldsymbol{x}_a,\boldsymbol{x}_\theta))} \\
&= \mathbb{E}_{\boldsymbol{y},\boldsymbol{y}',\xi} e^{\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} \frac{1}{n}\sum_{i=1}^n \xi_i(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\mathsf{w}_d(\boldsymbol{y}_i'-\boldsymbol{x}_a,\boldsymbol{x}_\theta))} \\
&\leq \mathbb{E}_{\boldsymbol{y},\boldsymbol{y}',\xi} e^{\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} |\frac{1}{n}\sum_{i=1}^n \xi_i(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\mathsf{w}_d(\boldsymbol{y}_i'-\boldsymbol{x}_a,\boldsymbol{x}_\theta))|} \\
&\leq \mathbb{E}_\xi\{\mathbb{E}_{\boldsymbol{y}}\left(e^{\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} |\frac{1}{n}\sum_{i=1}^n \xi_i(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})|}\right) \\
&\quad \times \mathbb{E}_{\boldsymbol{y}'}\left(e^{\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} |\frac{1}{n}\sum_{i=1}^n \xi_i(\mathsf{w}_d(\boldsymbol{y}_i'-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})|}\right)\} \\
&\leq \mathbb{E}_{\boldsymbol{y},\xi} e^{2\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} |\frac{1}{n}\sum_{i=1}^n \xi_i(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})|},
\end{aligned}$$

Note that to obtain Inequality (i) we have used Jensen's inequality. Also, ξ_i are i.i.d sequence of Rademacher variables. To simplify the final expression even further, we use the following lemma from Koltchinskii [2011]

Lemma A.9. Let $\mathcal{H} \in \mathbb{R}^n$ and let $\psi_i : \mathbb{R} \mapsto \mathbb{R}, i = 1, \cdots, n$ be functions such that $\psi_i(0) = 0$ and

$$|\psi_i(u) - \psi_i(v)| \leq |u - v| \in \mathbb{R}.$$

For all convex nondecreasing functions $\Psi : \mathbb{R}_+ \mapsto \mathbb{R}_+$,

$$\mathbb{E}\Psi(\frac{1}{2}\sup_{\boldsymbol{h}\in\mathcal{H}} |\sum_{i=1}^n \psi_i(h_i)\epsilon_i|) \leq \mathbb{E}\Psi(\sup_{\boldsymbol{h}\in\mathcal{H}} |\sum_{i=1}^n h_i\epsilon_i|),$$

where ϵ_i are i.i.d. Rademacher random variables.

Since $\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta)$ is a function of $\langle \boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta \rangle$ and

$$|\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta) - \mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a', \boldsymbol{x}_\theta')| \leq \frac{1}{2}|\langle \boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta \rangle - \langle \boldsymbol{y} - \boldsymbol{x}_a', \boldsymbol{x}_\theta' \rangle|,$$

letting $\Psi(x) = e^{2\lambda x}$ and $\psi_i(x) = \frac{2e^x}{e^x+e^{-x}} - 1$ with $h_i = \langle \boldsymbol{y}_i - \boldsymbol{x}_a, \boldsymbol{x}_\theta \rangle$ in Lemma A.9,

we have

$$\begin{aligned}
\mathbb{E}e^{\lambda Z_+} &\leq \mathbb{E}_{\boldsymbol{y},\xi} e^{2\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} |\frac{1}{n}\sum_{i=1}^n \xi_i(\mathsf{w}_d(\boldsymbol{y}_i,\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})|} \\
&\leq \mathbb{E}_{Y,\xi} e^{2\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} |\frac{1}{n}\sum_{i=1}^n \xi_i\langle \boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta\rangle|} \\
&\overset{(ii)}{=} \mathbb{E}_{\boldsymbol{y},\xi} e^{2\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} \frac{1}{n}\sum_{i=1}^n \xi_i\langle \boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta\rangle} \\
&\leq \mathbb{E}_{\boldsymbol{y},\xi} e^{2\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c} \frac{1}{n}\sum_{i=1}^n \xi_i\langle \boldsymbol{y}_i,\boldsymbol{x}_\theta\rangle} e^{2\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1}\langle \boldsymbol{x}_a,\boldsymbol{x}_\theta\rangle \frac{1}{n}\sum_{i=1}^n \xi_i} \\
&\leq \mathbb{E}_{\boldsymbol{y},\xi} e^{2\lambda c\|\frac{1}{n}\sum_{i=1}^n \xi_i\boldsymbol{y}_i\|} e^{2\lambda c|\frac{1}{n}\sum_{i=1}^n \xi_i|} \\
&\leq \big(\mathbb{E}_{\boldsymbol{y},\xi} e^{4\lambda c\|\frac{1}{n}\sum_{i=1}^n \xi_i\boldsymbol{y}_i\|}\big)^{1/2}\big(\mathbb{E}_{\xi} e^{4\lambda c|\frac{1}{n}\sum_{i=1}^n \xi_i|}\big)^{1/2} \\
&\leq \big(\underbrace{\mathbb{E}_{\boldsymbol{y}} e^{4\lambda c\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\|}}_{part\ 1}\big)^{1/2}\big(\underbrace{\mathbb{E}_{\xi} e^{4\lambda c|\frac{1}{n}\sum_{i=1}^n \xi_i|}}_{part\ 2}\big)^{1/2},
\end{aligned}$$

where last equality holds for the fact that the distribution of $\boldsymbol{y}_i$ is symmetric and equality (ii) holds for the fact that $\frac{1}{n}\sum_{i=1}^n \xi_i\langle \boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta\rangle$ is symmetric in terms of $\boldsymbol{x}_\theta$ and the constraints on $\boldsymbol{x}_\theta$ is symmetric.

For part 1, we use the notation $\{\boldsymbol{u}_j, j=1,\cdots,M\}$ for a 1/2-covering of the d-dimensional sphere, $Sp^d \triangleq \{\boldsymbol{v}\in\mathbb{R}^d, \|\boldsymbol{v}\|=1\}$. Note that, for all $\boldsymbol{v}',\boldsymbol{v}\in Sp^d$,

$$|\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{v}'\rangle - \frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{v}\rangle| \leq \|\boldsymbol{v}'-\boldsymbol{v}\| \sup_{\|\boldsymbol{u}\|=1}\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{u}\rangle,$$

therefore, we have for all $\boldsymbol{u}\in Sp^d$

$$\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{u}\rangle \leq \max_{j\in[M]}\{\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{u}_j\rangle\} + \|\boldsymbol{u}_j-\boldsymbol{u}\| \sup_{\|\boldsymbol{u}\|=1}\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{u}\rangle,$$

and hence

$$\frac{1}{n}\|\sum_{i=1}^n \boldsymbol{y}_i\| = \sup_{\|\boldsymbol{u}\|=1}\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{u}\rangle \leq 2\max_{j\in[M]}\{\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i,\boldsymbol{u}_j\rangle\}. \tag{A.82}$$

recall that $\boldsymbol{y}_i = \zeta_i\boldsymbol{\theta}^\star + \boldsymbol{\omega}_i$. Hence, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{y}} e^{\langle \boldsymbol{y}_i,\boldsymbol{u}_j\rangle} &= \mathbb{E}_{\zeta} e^{\zeta\langle \boldsymbol{\theta}^\star,\boldsymbol{u}_j\rangle}\mathbb{E}_{\boldsymbol{\omega}} e^{\langle \boldsymbol{\omega}_i,\boldsymbol{u}_j\rangle} \\
= \frac{1}{2}(e^{\langle \boldsymbol{\theta}^\star,\boldsymbol{u}_j\rangle} + e^{-\langle \boldsymbol{\theta}^\star,\boldsymbol{u}_j\rangle})e^{\frac{1}{2}} &\leq e^{\frac{\|\boldsymbol{\theta}^\star\|^2+1}{2}},
\end{aligned} \tag{A.83}$$

where last inequality holds because of

$$\frac{1}{2}(e^{\|\boldsymbol{\theta}^\star\|} + e^{-\|\boldsymbol{\theta}^\star\|}) \leq e^{\frac{\|\boldsymbol{\theta}^\star\|^2}{2}}.$$

Therefore we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{y},\xi} e^{4\lambda c\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\|} &= \mathbb{E}_{\boldsymbol{y},\xi} e^{4\lambda c \sup_{\|\boldsymbol{u}\|=1} \frac{1}{n}\sum_{i=1}^n \langle \boldsymbol{y}_i, \boldsymbol{u}\rangle} \\
&\leq \mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda c \max_{j\in[M]} \frac{1}{n}\sum_{i=1}^n \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle} \\
&\leq \sum_{j=1}^M \mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda c \frac{1}{n}\sum_{i=1}^n \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle} \\
&\leq e^{32\lambda^2 c^2 \frac{\|\boldsymbol{\theta}^\star\|^2+1}{n} + 2d}. \qquad \text{(A.84)}
\end{aligned}$$

For part 2, notice that $\frac{1}{n}\sum_{i=1}^n \xi_i$ is symmetric, we have

$$\begin{aligned}
\mathbb{E}_\xi e^{4\lambda c|\frac{1}{n}\sum_{i=1}^n \xi_i|} &\leq 2\mathbb{E}_\xi e^{4\lambda c \frac{1}{n}\sum_{i=1}^n \xi_i} \\
&\leq 2(\mathbb{E}_\xi e^{\frac{4\lambda c}{n}\xi})^n \\
&\leq e^{\frac{8\lambda^2 c^2}{n} + 1}. \qquad \text{(A.85)}
\end{aligned}$$

Therefore combining Equation (A.84) and Equation (A.85), we have

$$\begin{aligned}
\mathbb{E}e^{\lambda Z_+} &\leq e^{16\lambda^2 c^2 \frac{\|\boldsymbol{\theta}^\star\|^2+1}{n} + d} \times e^{\frac{4\lambda^2 c^2}{n} + \frac{1}{2}} \\
&\leq e^{16\lambda^2 c^2 \frac{\|\boldsymbol{\theta}^\star\|^2+2}{n} + d + \frac{1}{2}}.
\end{aligned}$$

Using Markov Inequality:

$$P(Z_+ > \epsilon) \leq \mathbb{E}e^{\lambda Z_+ - \lambda\epsilon}, \forall \epsilon, \lambda > 0,$$

choosing $\lambda = \frac{\epsilon n}{32c^2(\|\boldsymbol{\theta}^\star\|^2+2)}$, we have

$$\begin{aligned}
P(Z_+ > \epsilon) &\leq e^{\frac{16c^2\lambda^2(\|\boldsymbol{\theta}^\star\|^2+2)}{n} + d + \frac{1}{2} - \lambda\epsilon} \\
&= e^{-\frac{n\epsilon^2}{64c^2(\|\boldsymbol{\theta}^\star\|^2+2)} + d + \frac{1}{2}}.
\end{aligned}$$

Therefore

$$\left| \sup_{\substack{\|\boldsymbol{x}_\theta\|\leq c, \\ \|\boldsymbol{x}_a\|\leq 1}} \frac{1}{n}\sum_{i=1}^n \mathsf{w}_d(\boldsymbol{y}_i - \boldsymbol{x}_a, \boldsymbol{x}_\theta) - \mathbb{E}\mathsf{w}_d(\boldsymbol{y} - \boldsymbol{x}_a, \boldsymbol{x}_\theta) \right| \leq 8c(\|\boldsymbol{\theta}^\star\| + 2)\sqrt{\frac{d + 2 + \ln(1/\delta)}{n}},$$

with probability at least $1-\delta$.

For the last claim, we borrow a technique in the proof of corollary 2 in B.2 in Balakrishnan *et al.* [2017]. Let

$$Z = \sup_{\|\boldsymbol{x}_\theta\|\le c, \|\boldsymbol{x}_a\|\le 1} \|\frac{1}{n}\sum_{i=1}^{n}(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\boldsymbol{y}_i - \mathbb{E}(\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\boldsymbol{y}\|,$$

and

$$Z_{\boldsymbol{u}} = \sup_{\|\boldsymbol{x}_\theta\|\le c, \|\boldsymbol{x}_a\|\le 1} |\frac{1}{n}\sum_{i=1}^{n}(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2}))\langle\boldsymbol{y}_i,\boldsymbol{u}\rangle - \mathbb{E}(\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\langle\boldsymbol{y},\boldsymbol{u}\rangle|$$

we have

$$\begin{aligned}\mathbb{E}e^{\lambda Z} &= \mathbb{E}_Y e^{\lambda \sup_{\|\boldsymbol{u}\|=1} Z_{\boldsymbol{u}}} \le \mathbb{E}e^{2\lambda \max_{j\in[M]} Z_{\boldsymbol{u}_j}} \le \sum_{j=1}^{M}\mathbb{E}e^{2\lambda Z_{\boldsymbol{u}_j}} \\ &\le \sum_{j=1}^{M}\mathbb{E}_{\boldsymbol{y},\xi} e^{4\lambda \sup_{\|\boldsymbol{x}_\theta\|\le c,\|\boldsymbol{x}_a\|\le 1}\frac{1}{n}\sum_{i=1}^{n}\xi_i(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\langle\boldsymbol{y}_i,u_j\rangle},\end{aligned}$$

where ξ_i are i.i.d. sequence of Rademacher variables and the last inequality holds for standard symmetrization result for empirical process. Since

$$\begin{aligned}&|(2\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-1)\langle\boldsymbol{y}_i,\boldsymbol{u}_j\rangle - (2\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}'_a,\boldsymbol{x}'_\theta)-1)\langle\boldsymbol{y}_i,\boldsymbol{u}_j\rangle| \\ &\le |\langle\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta\rangle - \langle\boldsymbol{y}_i-\boldsymbol{x}'_a,\boldsymbol{x}'_\theta\rangle|\langle\boldsymbol{y}_i,\boldsymbol{u}_j\rangle,\end{aligned}$$

let $\Psi(x)=e^{2\lambda x}$ and $\psi_i(x)=(\frac{2e^x}{e^x-e^{-x}}-1)\langle\boldsymbol{y}_i,\boldsymbol{u}_j\rangle$ with $h_i=\langle\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta\rangle$ in Lemma

A.9, we have

$$
\begin{aligned}
\mathbb{E}e^{\lambda Z} &\leq \sum_{j=1}^{M} \mathbb{E}_{\boldsymbol{y},\xi} e^{4\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c, \|\boldsymbol{x}_a\|\leq 1} |\frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i - \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle|} \\
&\overset{iii}{=} \sum_{j=1}^{M} \mathbb{E}_{\boldsymbol{y},\xi} e^{4\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c, \|\boldsymbol{x}_a\|\leq 1} \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i - \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle} \\
&\leq \sum_{j=1}^{M} \mathbb{E}_{\boldsymbol{y},\xi} e^{4\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c} \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i, \boldsymbol{x}_\theta\rangle \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle} e^{4\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c, \|\boldsymbol{x}_a\|\leq 1} \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle} \\
&\leq \sum_{j=1}^{M} \Big(\mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c} \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i, \boldsymbol{x}_\theta\rangle \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle} \mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda \sup_{\|\boldsymbol{x}_\theta\|\leq c, \|\boldsymbol{x}_a\|\leq 1} \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle}\Big)^{\frac{1}{2}} \\
&\leq \sum_{j=1}^{M} \Big(\underbrace{\mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda c \|\frac{1}{n}\sum_{i=1}^{n} \xi_i \boldsymbol{y}_i \boldsymbol{y}_i^\top\|_{op}}}_{part1}\Big)^{\frac{1}{2}} \Big(\underbrace{\mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda c \frac{1}{n}|\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle|}}_{part2}\Big)^{\frac{1}{2}},
\end{aligned}
$$

where $\|\cdot\|_{op}$ is l_2-operator norm of a matrix(maximum singular value), equality (iii) holds for the fact that $\frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i - \boldsymbol{x}_a, \boldsymbol{x}_\theta\rangle \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle$ is symmetric in terms of $\boldsymbol{x}_\theta$ and constraints of $\boldsymbol{x}_\theta$ is symmetric. The correctness of the last inequality is shown in B.2 of Balakrishnan *et al.* [2017]. For part 1, as shown in B.2 of Balakrishnan *et al.* [2017] we have

$$
\mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda c \|\frac{1}{n}\sum_{i=1}^{n} \xi_i \boldsymbol{y}_i \boldsymbol{y}_i^\top\|_{op}} \leq \mathbb{E}_{\boldsymbol{y},\xi} e^{16\lambda c \max_{j'\in[M]} \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i, \boldsymbol{u}_{j'}\rangle^2} \tag{A.86}
$$

Recall that $\boldsymbol{y}_i = \zeta_i \boldsymbol{\theta}^\star + \boldsymbol{\omega}_i$ and Equation (A.83), we have

$$
\mathbb{E}_{\boldsymbol{y}} e^{\langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle} = \mathbb{E}_{\zeta} e^{\zeta \langle \boldsymbol{\theta}^\star, \boldsymbol{u}_j\rangle} \mathbb{E}_{\boldsymbol{\omega}} e^{\langle \boldsymbol{\omega}_i, \boldsymbol{u}_j\rangle} \leq e^{\frac{\|\boldsymbol{\theta}^\star\|^2+1}{2}}.
$$

Therefore

$$
\mathbb{E}e^{\lambda \xi \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle^2} \leq e^{\frac{(\|\boldsymbol{\theta}^\star\|^2+1)\lambda^2}{2}}, \text{ for small enough } \lambda.
$$

Therefore

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{y},\xi} e^{16\lambda c \max_{j'\in[M]} \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i, \boldsymbol{u}_{j'}\rangle^2} &\leq \sum_{j'=1}^{M} \mathbb{E}_{\boldsymbol{y},\xi} e^{16\lambda c \frac{1}{n}\sum_{i=1}^{n} \xi_i \langle \boldsymbol{y}_i, \boldsymbol{u}_{j'}\rangle^2} \\
&\leq e^{\frac{(\|\boldsymbol{\theta}^\star\|^2+1)(16c\lambda)^2}{2n} + 2d}.
\end{aligned}
$$

For part 2, since $\xi_i\langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle \overset{dist.}{=} \langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle$, using Equation (A.83), we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{y},\xi} e^{8\lambda c\frac{1}{n}|\sum_{i=1}^n \xi_i\langle \boldsymbol{y}_i,\boldsymbol{u}_j\rangle|} &= \mathbb{E}_{\boldsymbol{y}} e^{8\lambda c\frac{1}{n}|\sum_{i=1}^n \langle \boldsymbol{y}_i,\boldsymbol{u}_j\rangle|} \\ &\overset{iv}{\leq} 2\mathbb{E}_{\boldsymbol{y}} e^{8\lambda c\frac{1}{n}\sum_{i=1}^n \langle \boldsymbol{y}_i,\boldsymbol{u}_j\rangle} \\ &\leq e^{\frac{(\|\boldsymbol{\theta}^\star\|^2+1)(8c\lambda)^2}{2n}+1},\end{aligned}$$

where inequality (iv) holds for the fact that the distribution of $\frac{1}{n}\sum_{i=1}^n\langle \boldsymbol{y}_i, \boldsymbol{u}_j\rangle$ is symmetric. Therefore, combining part 1 and part 2, we have

$$\begin{aligned}\mathbb{E}e^{\lambda Z} &\leq \sum_{j=1}^M e^{\frac{(\|\boldsymbol{\theta}^\star\|^2+1)(16c\lambda)^2}{4}+d} e^{\frac{(\|\boldsymbol{\theta}^\star\|^2+1)(8c\lambda)^2}{4}+\frac{1}{2}} \\ &\leq e^{\frac{81(\|\boldsymbol{\theta}^\star\|^2+1)c^2\lambda^2}{n}+3d+\frac{1}{2}}.\end{aligned}$$

Using Markov Inequality:

$$P(Z>\epsilon) \leq \mathbb{E}_Y e^{\lambda Z-\lambda\epsilon}, \forall \epsilon, \lambda > 0,$$

choosing $\lambda = \frac{\epsilon n}{32c^2(\|\boldsymbol{\theta}^\star\|^2+2)}$, we have

$$\begin{aligned}P(Z>\epsilon) &\leq e^{\frac{81c^2\lambda^2(\|\boldsymbol{\theta}^\star\|^2+1)}{n}+3d+\frac{1}{2}-\lambda\epsilon} \\ &= e^{-\frac{n\epsilon^2}{324c^2(\|\boldsymbol{\theta}^\star\|^2+1)}+3d+\frac{1}{2}}.\end{aligned}$$

Therefore

$$\begin{aligned}&\sup_{\|\boldsymbol{x}_\theta\|\leq c,\|\boldsymbol{x}_a\|\leq 1} \|\tfrac{1}{n}\textstyle\sum_{i=1}^n(\mathsf{w}_d(\boldsymbol{y}_i-\boldsymbol{x}_a,\boldsymbol{x}_\theta)-\frac{1}{2})\boldsymbol{y}_i - \mathbb{E}\mathsf{w}_d(\boldsymbol{y}-\boldsymbol{x}_a,\boldsymbol{x}_\theta)\boldsymbol{y}\| \\ &\leq 36c(\|\boldsymbol{\theta}^\star\|+2)\sqrt{\tfrac{d+2+\ln(1/\delta)}{n}},\end{aligned}$$

with probability at least $1-\delta$. □

A.2.1.3 Proof of Lemma 2.21

Since $\bar{\boldsymbol{a}}^{\langle t+1\rangle}$ and $\bar{\boldsymbol{\theta}}^{\langle t+1\rangle}$ are the result of first iteration based on initialization $(\hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})$ in Population EM model where initialization $(\hat{\boldsymbol{a}}^{\langle t\rangle}, \hat{\boldsymbol{\theta}}^{\langle t\rangle})$ satisfying the corresponding condition mentioned in the lemma. Hence to prove the lemma holds for all $t \geq 0$, it is sufficient to prove that for any initialization $(\boldsymbol{a}^{\langle 0\rangle}, \boldsymbol{\theta}^{\langle 0\rangle})$ satisfying the same condition, we have $\|\boldsymbol{a}^{\langle 1\rangle}\| \leq \kappa_a\|\boldsymbol{a}^{\langle 0\rangle}\|$ for Equation (2.109) and $\|\boldsymbol{\theta}^{\langle 1\rangle} - \boldsymbol{\theta}^\star\| \leq \kappa_\theta\|\boldsymbol{\theta}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\| +$

$\sqrt{c_\theta\|\boldsymbol{a}^{\langle 0\rangle}\|}$ for Equation (2.110). To achieve this goal, we use the sequence of the coordinate systems $\mathcal{A}$. We first prove the first claim:

$$\|\boldsymbol{a}^{\langle 1\rangle}\| \leq \kappa_a \|\boldsymbol{a}^{\langle 0\rangle}\|. \tag{A.87}$$

If $\boldsymbol{a}^{\langle 0\rangle} = \mathbf{0}$, we immediately have Equation (A.87) holds. If $\boldsymbol{a}^{\langle 0\rangle} \neq \mathbf{0}$, because of Lemma 2.4, we assume $\langle \boldsymbol{a}^{\langle 0\rangle}, \boldsymbol{\theta}^{\langle 0\rangle}\rangle > 0$ without loss of generality, thus $a_1^{\langle 0\rangle} > 0$. Since $\boldsymbol{\theta}^{\langle 1\rangle}(1-2\mathsf{p}^{\langle 1\rangle}) = \boldsymbol{a}^{\langle 1\rangle}$, we know they are in the same direction, thus the angle between $\boldsymbol{a}^{\langle 1\rangle}$ and $\boldsymbol{\theta}^{\langle 0\rangle}$ is the same angle between $\boldsymbol{\theta}^{\langle 1\rangle}$ and $\boldsymbol{\theta}^{\langle 0\rangle}$, i.e., $\alpha^{\langle 1\rangle}$. Furthermore, according to the proof of Lemma 2.8, we have $\alpha^{\langle 1\rangle} \leq \beta^{\langle 0\rangle}$. Hence we have

$$\|\boldsymbol{a}^{\langle 1\rangle}\| = \frac{a_{\langle 0\rangle,1}^{\langle 1\rangle}}{\cos\alpha^{\langle 1\rangle}} \leq \frac{a_{\langle 0\rangle,1}^{\langle 1\rangle}}{\cos\beta^{\langle 0\rangle}}. \tag{A.88}$$

Therefore, we need to bound $a_{\langle 0\rangle,1}^{\langle 1\rangle}$ and $\frac{1}{\cos\beta^{\langle 0\rangle}}$. According to Equation (2.53) and Lemma 2.11 we have,

$$a_{\langle 0\rangle,1}^{\langle 1\rangle} = \frac{g_\gamma(a_1^{\langle 0\rangle}, \|\boldsymbol{\theta}^{\langle 0\rangle}\|, \theta^\star_{\langle 0\rangle,1})(1-2g_p(a_1^{\langle 0\rangle}, \|\boldsymbol{\theta}^{\langle 0\rangle}\|, \theta^\star_{\langle 0\rangle,1}))}{g_p(a_1^{\langle 0\rangle}, \|\boldsymbol{\theta}^{\langle 0\rangle}\|, \theta^\star_{\langle 0\rangle,1})(1-g_p(a_1^{\langle 0\rangle}, \|\boldsymbol{\theta}^{\langle 0\rangle}\|, \theta^\star_{\langle 0\rangle,1}))} \leq \kappa_a' a_1^{\langle 0\rangle} \leq \kappa_a' \|\boldsymbol{a}^{\langle 0\rangle}\|,$$

where $\kappa_a' \in (0,1)$ is a continuous function of $\theta^\star_{\langle 0\rangle,1} > 0$. Since the condition of $\|\boldsymbol{\theta}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\| \leq \frac{1}{2}\|\boldsymbol{\theta}^\star\|$ implies that $\theta^\star_{\langle 0\rangle,1} \geq \frac{\sqrt{3}}{2}\|\boldsymbol{\theta}^\star\| > 0$, we have

$$\kappa_a \triangleq \sup_{\theta^\star_{\langle 0\rangle,1} \in [\frac{\sqrt{3}}{2}\|\boldsymbol{\theta}^\star\|, \|\boldsymbol{\theta}^\star\|]} \sqrt{\kappa_a'(\theta^\star_{\langle 0\rangle,1})} \in (0,1),$$

and κ_a only depends on $\boldsymbol{\theta}^\star$. Now for $\frac{1}{\cos\beta^{\langle 0\rangle}}$, by the condition of $\|\boldsymbol{\theta}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\| \leq \sqrt{1-(\kappa_a)^2}\|\boldsymbol{\theta}^\star\|$, we have $\cos\beta^{\langle 0\rangle} \geq \kappa_a$. Hence, combining the two parts in Equation (A.88), we have

$$\|\boldsymbol{a}^{\langle 1\rangle}\| \leq \frac{a_{\langle 0\rangle,1}^{\langle 1\rangle}}{\cos\beta^{\langle 0\rangle}} \leq \frac{\kappa_a'(\theta^\star_{\langle 0\rangle,1})}{\kappa_a}\|\boldsymbol{a}^{\langle 0\rangle}\| \leq \kappa_a \|\boldsymbol{a}^{\langle 0\rangle}\|.$$

Hence Equation (A.87) holds. Next we prove the second claim:

$$\|\boldsymbol{\theta}^{\langle 1\rangle} - \boldsymbol{\theta}^\star\| \leq \kappa_\theta \|\boldsymbol{\theta}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_\theta \|\boldsymbol{a}^{\langle 0\rangle}\|}.$$

According to Lemma 2.12 we can conclude there exists $\delta_a' \in (0,1)$, $\kappa_\theta \in (0,1)$ and $c_\theta > 0$ such that that if $\|\boldsymbol{a}^{\langle 0\rangle}\| \leq \delta_a'$, then

$$\|\boldsymbol{\theta}^{\langle 1\rangle} - \boldsymbol{\theta}^\star\| \;\; \leq \;\; \sqrt{\kappa_b^2\|\boldsymbol{\theta}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\|^2 + c_b\|\boldsymbol{a}^{\langle 0\rangle}\|} \leq \kappa_\theta\|\boldsymbol{\theta}^{\langle 0\rangle} - \boldsymbol{\theta}^\star\| + \sqrt{c_b\|\boldsymbol{a}^{\langle 0\rangle}\|},$$

where δ_a', κ_θ and c_θ only depend on $U_a = 1, L_\theta = \frac{1}{2}\|\boldsymbol{\theta}^\star\|, U_\theta = \frac{3}{2}\|\boldsymbol{\theta}^\star\|, L_{\theta^\star} = \frac{\sqrt{3}}{2}\|\boldsymbol{\theta}^\star\|$ and $\|\boldsymbol{\theta}^\star\|$. Hence δ_a', κ_θ and c_θ only depend on $\boldsymbol{\theta}^\star$. This completes the proof.

A.2.2 Proofs omitted in Sections 2.6

A.2.2.1 Proof of Theorem 2.8

The maximum log-likelihood objective for Population EM of Model 4 is the following optimization problem:

$$\max_{\boldsymbol{\theta}\in\mathbb{R}^d, w_1\in[0,1]} \mathbb{E}_{\boldsymbol{y}\sim f^*} \log\left(w_1 e^{-\frac{\|\boldsymbol{y}-\boldsymbol{\theta}\|^2}{2}} + w_2 e^{-\frac{\|\boldsymbol{y}+\boldsymbol{\theta}\|^2}{2}}\right). \tag{A.89}$$

Due to the symmetric property of the landscape, without loss of generality, we assume $w_1^\star > 0.5$.

Note that the first order stationary points of above optimization problem should satisfy the following equation.

$$\mathbb{E}_{\boldsymbol{y}\sim f^*}\left[\frac{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle} - w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle} + w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}\boldsymbol{y}\right] - \boldsymbol{\theta} \;\; = \;\; \mathbf{0}, \tag{A.90}$$

$$\mathbb{E}_{\boldsymbol{y}\sim f^*}\left[\frac{e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle} - e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}{w_1 e^{\langle \boldsymbol{y},\boldsymbol{\theta}\rangle} + w_2 e^{-\langle \boldsymbol{y},\boldsymbol{\theta}\rangle}}\right] \;\; = \;\; 0. \tag{A.91}$$

We first consider the two trivial cases when $w_1 = 1$ and $w_1 = 0$. Suppose $w_1 = 1$, then from Equation (A.90), we have $\boldsymbol{\theta} = (w_1^\star - w_2^\star)\boldsymbol{\theta}^\star$. Now plug it in Equation (A.91), we have the following equation holds

$$\int \left(1 - e^{-2(w_1^\star - w_2^\star)y\|\boldsymbol{\theta}^\star\|}\right)\left(w_1^\star\phi(y - \|\boldsymbol{\theta}^\star\|) + w_2^\star\phi(y + \|\boldsymbol{\theta}^\star\|)\right)\mathrm{d}y \;\; = \;\; 0,$$

which is equivalent to

$$1 - w_1^\star e^{-4w_2^\star(w_1^\star - w_2^\star)\|\boldsymbol{\theta}^\star\|^2} - w_2^\star e^{4w_1^\star(w_1^\star - w_2^\star)\|\boldsymbol{\theta}^\star\|^2} \;\; = \;\; 0.$$

Taking the derivative with respect to $\|\boldsymbol{\theta}^\star\|$, it is straightforward to show that when

$w_1^\star > 0.5$, the LHS is a strictly decreasing function of $\|\boldsymbol{\theta}^\star\|$ and achieves its maximum 0 at $\|\boldsymbol{\theta}^\star\| = 0$. Hence, it contradicts the RHS of the equation and therefore Equation (A.90) and Equation (A.91) can not hold simultaneously for $w_1 = 1$. Hence, there is no first order stationary point for the case $w_1 = 1$. Further, based on above calculation, it is straightforward to show that there is no local optimum in general on the boundary $w_1 = 1$ and similarly for $w_1 = 0$.

Now we restrict $w_1 \in (0, 1)$. Then it is straightforward to show that every first order stationary point of the optimization in Equation (A.89) should be a fixed point for population-EM$_2$. From the proof of Theorem 2.4, we know the two global maxima $(\boldsymbol{\theta}^\star, w_1^\star)$ and $(-\boldsymbol{\theta}^\star, w_2^\star)$ are the only fixed points of population-EM$_2$ in the following region:

$$\underbrace{\{(\boldsymbol{\theta}, w_1) | w_1 \in [0.5, 1), \langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star \rangle > 0\}}_{\text{Area}_1} \bigcup \underbrace{\{(\boldsymbol{\theta}, w_1) | w_1 \in (0, 0.5], \langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star \rangle < 0\}}_{\text{Area}_2}$$

Furthermore, for any fixed point lies in the hyperplane $\mathcal{H} : \langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star \rangle = 0$, it is clear that its corresponding w_1 should be 0.5. Further, since $\langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star \rangle = 0$, from Equation (A.90), it is clear that $\boldsymbol{\theta}$ should satisfy the following equation

$$\int \frac{e^{y\|\boldsymbol{\theta}\|} - e^{-y\|\boldsymbol{\theta}\|}}{e^{y\|\boldsymbol{\theta}\|} + e^{-y\|\boldsymbol{\theta}\|}} y \phi(y) \mathrm{d}y \quad = \quad \|\boldsymbol{\theta}\|.$$

Since the derivative with respect to $\|\boldsymbol{\theta}\|$ of the LHS is in $(0, 1)$ for $\|\boldsymbol{\theta}\| > 0$, it is clear that $\|\boldsymbol{\theta}\| = 0$ is the only solution for the equation and therefore, $(\boldsymbol{\theta}, w_1) = (\mathbf{0}, \frac{1}{2})$ is the only fixed point in the hyperplane $\mathcal{H}$. Furthermore, the Hessian of the log-likelihood in Equation (A.89) at $(\boldsymbol{\theta}, w_1) = (\mathbf{0}, \frac{1}{2})$ is the following matrix.

$$\begin{bmatrix} \boldsymbol{\theta}^\star (\boldsymbol{\theta}^\star)^\top & 2(w_1^\star - w_2^\star)\boldsymbol{\theta}^\star \\ 2(w_1^\star - w_2^\star)(\boldsymbol{\theta}^\star)^\top & 0 \end{bmatrix} \tag{A.92}$$

It is clear that it has a positive eigenvalue, a negative eigenvalue and therefore $(\mathbf{0}, \frac{1}{2})$ is a saddle point.

Finally, we will show there is no fixed point in the rest of the region in $\mathbb{R}^2 \times [0, 1]$, i.e.,

$$\underbrace{\{(\boldsymbol{\theta}, w_1) | w_1 \in (0, 0.5), \langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star \rangle > 0\}}_{\text{Area}_3} \bigcup \underbrace{\{(\boldsymbol{\theta}, w_1) | w_1 \in (0.5, 1), \langle \boldsymbol{\theta}, \boldsymbol{\theta}^\star \rangle < 0\}}_{\text{Area}_4}$$

Due to the symmetric property, we will just prove the result for Area$_3$. Note that, by

Lemma 2.15 and the fact that

$$g_w(\theta, 0.5) \leqslant 0.5, \quad \forall \theta \lessgtr 0. \tag{A.93}$$

We know for all $w_1 \in (0, 0.5)$,

$$\begin{aligned} 0 &< g_w(\|\boldsymbol{\theta}\|, w_1; \theta_\|, w_1^\star) - w_1 \\ &= w_1 w_2 \int \left[\frac{e^{y\|\boldsymbol{\theta}\|} - e^{-y\|\boldsymbol{\theta}\|}}{w_1 e^{y\|\boldsymbol{\theta}\|} + w_2 e^{-y\|\boldsymbol{\theta}\|}} \right] \left(w_1^* \phi(y - \theta_\|) + w_2^* \phi(y + \theta_\|) \right) \mathrm{d}y \\ &= w_1 w_2 \cdot \mathbb{E}_{\boldsymbol{y} \sim f^*} \left[\frac{e^{\langle \boldsymbol{y}, \boldsymbol{\theta} \rangle} - e^{-\langle \boldsymbol{y}, \boldsymbol{\theta} \rangle}}{w_1 e^{\langle \boldsymbol{y}, \boldsymbol{\theta} \rangle} + w_2 e^{-\langle \boldsymbol{y}, \boldsymbol{\theta} \rangle}} \right], \end{aligned}$$

where $\theta_\| = \langle \boldsymbol{\theta}^\star, \boldsymbol{\theta} \rangle / \|\boldsymbol{\theta}\|$. Hence, there is no solution for Equation (A.91) in Area_3. This completes the proof of this theorem.

A.3 Proofs of asymptotics of AMP.A in complex-valued case

A.3.1 Simplifications of SE maps

A.3.1.1 Auxiliary Results

Here we collect some auxiliary results that will be used in the simplification of the state evolution equation.

Lemma A.10. The following identities hold for any $a \in \mathbb{R}$ and $b \in \mathbb{R}_+$:

$$\int_0^{2\pi} \int_0^{\infty} r \cos\theta \exp\left(-\frac{r^2 - 2ar\cos\theta}{b} \right) \mathrm{d}r\mathrm{d}\theta = 2a\sqrt{b}\sqrt{\pi} \int_0^{\frac{\pi}{2}} \cos^2\theta \exp\left(\frac{a^2\cos^2\theta}{b} \right) \mathrm{d}\theta, \tag{A.94a}$$

$$\int_0^{2\pi} \int_0^{\infty} r \sin\theta \exp\left(-\frac{r^2 - 2ar\cos\theta}{b} \right) \mathrm{d}r\mathrm{d}\theta = 0. \tag{A.94b}$$

Proof. We first consider Equation (A.94a):

$$
\begin{aligned}
&\int_0^{2\pi}\int_0^{\infty} r\cos\theta \exp\left(-\frac{r^2-2a\cdot r\cos\theta}{b}\right)\mathrm{d}\theta\mathrm{d}r \\
&= \int_0^{2\pi} \cos\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta \int_0^{\infty} r\exp\left(-\frac{(r-a\cos\theta)^2}{b}\right)\mathrm{d}r \\
&\overset{\text{(i)}}{=} \int_0^{2\pi} \cos\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right)\left[\frac{1}{2}b\exp\left(\frac{-a^2\cos^2\theta}{b}\right) + a\cos\theta\sqrt{b\pi}\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\right]\mathrm{d}\theta \\
&= \int_0^{2\pi} \frac{1}{2}b\cos\theta\mathrm{d}\theta + \int_0^{2\pi} a\cos^2\theta\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\theta}{b}\right)\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\mathrm{d}\theta \\
&\overset{\text{(ii)}}{=} \int_0^{\pi} a\cos^2\theta\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\theta}{b}\right)\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\mathrm{d}\theta \\
&\quad + \int_0^{\pi} a\cos^2\hat{\theta}\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\hat{\theta}}{b}\right)\Phi\left(-\frac{\sqrt{2}a\cos\hat{\theta}}{\sqrt{b}}\right)\mathrm{d}\hat{\theta} \\
&= \int_0^{\pi} a\cos^2\theta\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\theta}{b}\right)\left[\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right) + \Phi\left(-\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\right]\mathrm{d}\theta \\
&\overset{\text{(iii)}}{=} a\sqrt{b\pi}\int_0^{\pi}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta \\
&\overset{\text{(iv)}}{=} 2a\sqrt{b\pi}\int_0^{\frac{\pi}{2}}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta,
\end{aligned}
\tag{A.95}
$$

where Equation (i) is from the integral ($\Phi(x)$ denotes the CDF of the standard Gaussian distribution):

$$
\int_0^{\infty} r\exp\left(-\frac{(r-m)^2}{v}\right)\mathrm{d}r = \frac{1}{2}b\exp\left(\frac{-m^2}{v}\right) + m\sqrt{v\pi}\Phi\left(\frac{\sqrt{2}m}{\sqrt{v}}\right), \quad \forall m\in\mathbb{R}, v\in\mathbb{R}_+,
$$

Equation (ii) is from the variable change $\hat{\theta} = \theta - \pi$, Equation (iii) is from the fact

that $\Phi(x) + \Phi(-x) = 1$, and Equation (iv) is from

$$\begin{aligned}
&\int_0^{\pi} \cos^2\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right) \mathrm{d}\theta \\
&= \int_0^{\frac{\pi}{2}} \cos^2\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right) \mathrm{d}\theta + \int_{\frac{\pi}{2}}^{\pi} \cos^2\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right) \mathrm{d}\theta \\
&= \int_0^{\frac{\pi}{2}} \cos^2\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right) \mathrm{d}\theta + \int_{\frac{\pi}{2}}^{0} \cos^2\hat{\theta} \exp\left(\frac{a^2\hat{\cos}^2\theta}{b}\right) (-\mathrm{d}\hat{\theta}) \quad (\hat{\theta} = \pi - \theta) \\
&= 2\int_0^{\frac{\pi}{2}} \cos^2\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right) \mathrm{d}\theta.
\end{aligned}$$

The identity in Equation (A.94b) can be derived based on similar calculations:

$$\begin{aligned}
\int_0^{2\pi}\int_0^{\infty} r\sin\theta \exp\left(-\frac{r^2 - 2b\cdot r\cos\theta}{b}\right) \mathrm{d}\theta\mathrm{d}r &= a\sqrt{b\pi}\int_0^{\pi} \frac{1}{2}\sin 2\theta \exp\left(\frac{a^2\cos^2\theta}{b}\right) \mathrm{d}\theta \\
&= 0.
\end{aligned}$$

□

Lemma A.11. Let $\tilde{Z} \sim \mathcal{N}(0,1)$ be a standard Gaussian random variable. Then, for any $x \in \mathbb{R}$, the following identities hold:

$$\begin{aligned}
\mathbb{E}\left[|\tilde{Z}| \cdot \phi\left(x|\tilde{Z}|\right)\right] &= \frac{1}{\pi}\frac{1}{1+x^2}, \\
\mathbb{E}\left[\Phi\left(x|\tilde{Z}|\right)\right] &= \frac{1}{\pi}\arctan(x) + \frac{1}{2}, \\
\mathbb{E}\left[\tilde{Z}^2 \cdot \Phi\left(x|\tilde{Z}|\right)\right] &= \frac{1}{\pi}\arctan(x) + \frac{1}{2} + \frac{1}{\pi}\frac{x}{1+x^2},
\end{aligned} \tag{A.96}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, PDF and CDF functions of the standard Gaussian distribution.

Proof. Consider the first identity:

$$\begin{aligned}\mathbb{E}\left[|\tilde{Z}|\cdot\phi\left(x|\tilde{Z}|\right)\right] &= \int_{-\infty}^{\infty}|z|\phi(x\,|z|)\phi(z)\mathrm{d}z \\ &\overset{\text{(i)}}{=} 2\int_{0}^{\infty} z\phi(x\,z)\phi(z)\mathrm{d}z \\ &\overset{\text{(ii)}}{=} \frac{1}{\pi}\int_{0}^{\infty} z\exp\left[-(1+x^2)\frac{z^2}{2}\right]\mathrm{d}z \\ &= \frac{1}{\pi}\frac{1}{1+x^2},\end{aligned} \tag{A.97}$$

where Equation (i) is from the symmetry of ϕ and Equation (ii) is from the definition $\phi(x) = 1/\sqrt{2\pi}e^{-x^2/2}$. Further,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{E}\left[\Phi\left(x|\tilde{Z}|\right)\right] &= \frac{\mathrm{d}}{\mathrm{d}x}\int_{-\infty}^{\infty}\Phi(x\,|z|)\phi(z)\mathrm{d}z = \frac{\mathrm{d}}{\mathrm{d}x}\int_{0}^{\infty}2\Phi(x\,z)\phi(z)\mathrm{d}z \\ &= \int_{0}^{\infty}2\frac{\mathrm{d}}{\mathrm{d}x}\Phi(x\,z)\phi(z)\mathrm{d}z = \int_{0}^{\infty}2z\phi(x\,z)\phi(z)\mathrm{d}z \\ &= \frac{1}{\pi}\frac{1}{1+x^2},\end{aligned} \tag{A.98}$$

where the last equality is from Equation (A.97). Hence,

$$\mathbb{E}\left[\Phi\left(x|\tilde{Z}|\right)\right] = \int_{-\infty}^{x}\frac{1}{\pi}\frac{1}{1+t^2}\mathrm{d}t = \frac{1}{\pi}\arctan(x)+\frac{1}{2}. \tag{A.99}$$

Finally, the third identity in Equation (A.96) can be derived as follows:

$$\begin{aligned}\mathbb{E}\left[\tilde{Z}^2\cdot\Phi\left(x|\tilde{Z}|\right)\right] &= \int_{-\infty}^{\infty}z^2\Phi(x\,|z|)\phi(z)\mathrm{d}z \\ &= \int_{0}^{\infty}z^2\Phi(xz)\phi(z)\mathrm{d}z \\ &\overset{\text{(i)}}{=} -2\int_{0}^{\infty}z\Phi(xz)\mathrm{d}\phi(z) \\ &= -2\left\{z\Phi(xz)\phi(z)\big|_0^{\infty} - \int_{0}^{\infty}\phi(z)\left[\Phi(xz)+xz\phi(xz)\mathrm{d}z\right]\right\} \\ &= 2\int_{0}^{\infty}\phi(z)\Phi(xz)\mathrm{d}z + x\cdot 2\int_{0}^{\infty}z\phi(xz)\phi(z)\mathrm{d}z \\ &\overset{\text{(ii)}}{=} \frac{1}{\pi}\arctan(x)+\frac{1}{2}+\frac{1}{\pi}\frac{x}{1+x^2},\end{aligned} \tag{A.100}$$

where Equation (i) is from the identity $\phi'(z) = z\phi(z)$ and Equation (ii) from our previously derived identities in Equation (A.97) and Equation (A.99). □

A.3.1.2 Complex-valued AMP.A

From Definition 3, the SE equations are given by

$$\begin{aligned}
\psi_1(\alpha, \sigma^2) &= 2 \cdot \mathbb{E}\left[\partial_z g(p, Y)\right] \\
&= \mathbb{E}\left[\frac{\bar{Z}P}{|Z|\,|P|}\right], \\
\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) &= 4 \cdot \mathbb{E}\left[|g(P, Y)|^2\right] \\
&= 4 \cdot \mathbb{E}\left[(|Z| - |P| + W)^2\right] \\
&= \underbrace{4 \cdot \mathbb{E}\left[(|Z| - |P|)^2\right]}_{\psi_2(\alpha, \sigma^2; \delta)} + 4\sigma_w^2.
\end{aligned} \tag{A.101}$$

In the above, $Z \sim \mathcal{CN}(0, 1/\delta)$, $P = \alpha Z + \sigma B$ where $B \sim \mathcal{CN}(0, 1/\delta)$ is independent of Z, and $Y = |Z| + W$ where $W \sim \mathcal{CN}(0, \sigma_w^2)$ independent of both Z and B. We first consider a special case $\sigma^2 = 0$ ($\alpha \neq 0$). When $\sigma = 0$, we have $P = \alpha Z + \sigma B = \alpha Z$, and therefore

$$\begin{aligned}
\psi_1(\alpha, 0) &= \mathbb{E}\left[\frac{\alpha \bar{Z} Z}{\alpha |Z|\,|Z|}\right] = 1, \\
\psi_2(\alpha, 0; \delta, \sigma_w^2) &= 4 \cdot \mathbb{E}\left[(|Z| - |\alpha Z|)^2\right] + 4\sigma_w^2 = \frac{4}{\delta}\left(1 - |\alpha|\right)^2 + 4\sigma_w^2.
\end{aligned}$$

We next turn to the general case where $\sigma^2 \neq 0$. Later, we will see that our formulas derived for positive σ^2 covers the special case $\sigma^2 = 0$ as well. Lemma A.12 can simplify our derivations.

Lemma A.12. ψ_1 and ψ_2 in Equation (A.101) have the following properties (for any $\alpha \in \mathbb{C}\backslash 0$ and $\sigma^2 \geq 0$):

(i) $\psi_1(\alpha, \sigma^2) = \psi_1(|\alpha|, \sigma^2) \cdot e^{\mathrm{i}\theta_\alpha}$, with $e^{\mathrm{i}\theta_\alpha}$ being the phase of α;

(ii) $\psi_2(\alpha, \sigma^2; \delta) = \psi_2(|\alpha|, \sigma^2; \delta)$.

Proof. Note that for ψ_1 and ψ_2 defined in Equation (A.101), we have

$$P|Z \sim \mathcal{CN}(\alpha Z, \sigma^2/\delta).$$

Consider the random variable $\tilde{P} \triangleq P \cdot e^{-\mathrm{i}\theta_\alpha}$. Based on the rotational invariance of circularly-symmetric Gaussian, we have $\tilde{P}|Z \sim \mathcal{CN}(|\alpha|Z; \sigma^2/\delta)$. Hence,

$$\psi_1(\alpha, \sigma^2) = \mathbb{E}\left[\frac{\bar{Z}P}{|Z|\,|P|}\right] = e^{\mathrm{i}\theta_\alpha} \cdot \mathbb{E}\left[\frac{\bar{Z}\tilde{P}}{|Z|\,|\tilde{P}|}\right] = e^{\mathrm{i}\theta_\alpha} \cdot \psi_1(|\alpha|, \sigma^2).$$

The proof of $\psi_2(\alpha, \sigma^2; \delta) = \psi_2(|\alpha|, \sigma^2; \delta)$ follows from a similar argument: the joint distribution of $|Z|$ and $|P|$ does not depend on θ_α, and thus $\psi_2(\alpha, \sigma^2) = 4\mathbb{E}\left[(|Z| - |P|)^2\right]$ does not depend on θ_α. □

Note that Lemma A.12 also holds for $\alpha = 0$ if we define $\angle 0 = 0$.

Remark A.1. In the following, we will derive ψ_1 and ψ_2 for the case where α *is real and nonnegative*. The results for complex-valued α can be easily derived from those for nonnegative α, based on Lemma A.12.

We can also write ψ_1 as

$$\psi_1(\alpha, \sigma^2) = \mathbb{E}\left[\frac{\bar{Z}P}{|Z|\,|P|}\right] = \mathbb{E}[e^{\mathrm{i}(\theta_p - \theta_z)}].$$

Note that $\theta_p - \theta_z$ is the phase of an auxiliary variable $\hat{P} \triangleq e^{-\mathrm{i}\theta_z}P = \alpha|Z| + \sigma e^{-\mathrm{i}\theta_z}B$. Further, from the rotational invariance, conditioned on $|Z|$, $\hat{P}$ is distributed as $\hat{P} \sim \mathcal{CN}(\alpha|Z|, \sigma^2/\delta)$. Hence, the expectation of its phase can be calculated as

$$\begin{aligned}
\mathbb{E}\left[e^{\mathrm{i}(\theta_p - \theta_z)}\middle|\,|Z|\right] &= \int_0^{2\pi}\int_0^{\infty} e^{\mathrm{i}\theta} \cdot \frac{1}{\pi\sigma^2/\delta} \exp\left(-\frac{\left|re^{\mathrm{i}\theta} - \alpha|Z|\right|^2}{\sigma^2/\delta}\right) \cdot r\mathrm{d}r\mathrm{d}\theta \\
&= \frac{1}{\pi\sigma^2/\delta} \exp\left(-\frac{\alpha^2|Z|^2}{\sigma^2/\delta}\right) \cdot \int_0^{2\pi}\int_0^{\infty} re^{\mathrm{i}\theta} \cdot \frac{1}{\pi\sigma^2/\delta} e^{-\frac{r^2 - 2\alpha|Z|\cos\theta r}{\sigma^2/\delta}} \mathrm{d}r\mathrm{d}\theta \\
&= \mathrm{i}\frac{1}{\pi\sigma^2/\delta} \exp\left(-\frac{\alpha^2|Z|^2}{\sigma^2/\delta}\right) \cdot \int_0^{2\pi}\int_0^{\infty} r\sin\theta \cdot \frac{1}{\pi\sigma^2/\delta} e^{-\frac{r^2 - 2\alpha|Z|\cos\theta r}{\sigma^2/\delta}} \mathrm{d}r\mathrm{d}\theta \\
&\quad + \frac{1}{\pi\sigma^2/\delta} \exp\left(-\frac{\alpha^2|Z|^2}{\sigma^2/\delta}\right) \cdot \int_0^{2\pi}\int_0^{\infty} r\cos\theta \cdot \frac{1}{\pi\sigma^2/\delta} e^{-\frac{r^2 - 2\alpha|Z|\cos\theta r}{\sigma^2/\delta}} \mathrm{d}r\mathrm{d}\theta \\
&= 2\int_0^{\frac{\pi}{2}} \frac{\alpha|Z|}{\sqrt{\pi}\sqrt{\sigma^2/\delta}} \cos^2\theta \exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right) \mathrm{d}\theta, \qquad \text{(A.102)}
\end{aligned}$$

where the last step follow the following two identities together with some straightfor-

ward manipulations:

$$\begin{aligned}\int_0^{2\pi}\int_0^{\infty} r\cos\theta e^{-\frac{r^2-2\alpha|Z|\cos\theta r}{\sigma^2/\delta}}\mathrm{d}r\mathrm{d}\theta &= \frac{2\alpha\sigma\sqrt{\pi}}{\sqrt{\delta}}\int_0^{\frac{\pi}{2}}\cos^2\theta e^{\frac{\alpha^2|Z|^2\cos^2\theta}{\sigma^2/\delta}}\mathrm{d}\theta,\\ \int_0^{2\pi}\int_0^{\infty} r\sin\theta e^{-\frac{r^2-2\alpha|Z|\cos\theta r}{\sigma^2/\delta}}\mathrm{d}r\mathrm{d}\theta &= 0.\end{aligned}$$

The above identities are proved in Lemma A.10 in Appendix A.3.1.1. Using Equation (A.102) and noting that $Z\sim\mathcal{CN}(0,1/\delta)$, we further average our result over $|Z|$:

$$\begin{aligned}\mathbb{E}\left[e^{\mathrm{i}(\theta_p-\theta_z)}\right] &= \mathbb{E}\left\{2\int_0^{\frac{\pi}{2}}\frac{\alpha|Z|}{\sqrt{\pi}\sqrt{\sigma^2/\delta}}\cos^2\theta\exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta\right\}\\ &\overset{\text{(i)}}{=}\int_0^{\infty}2\delta r\exp\left(-\delta r^2\right)\cdot\left(2\int_0^{\frac{\pi}{2}}\frac{\alpha r}{\sqrt{\pi}\sqrt{\sigma^2/\delta}}\cos^2\theta\exp\left(-\frac{\alpha^2r^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta\right)\mathrm{d}r\\ &= \frac{4\alpha\delta^{3/2}}{\sqrt{\pi}\sigma}\int_0^{\frac{\pi}{2}}\cos^2\theta\mathrm{d}\theta\int_0^{\infty}r^2\exp\left(-\delta\left(1+\frac{\alpha^2\sin^2\theta}{\sigma^2}\right)r^2\right)\mathrm{d}r\\ &\overset{\text{(ii)}}{=}\frac{\alpha}{\sigma}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2\sin^2\theta}{\sigma^2}\right)^{-\frac{3}{2}}\mathrm{d}\theta\\ &\overset{\text{(iii)}}{=}\frac{\alpha}{\sigma}\int_0^{\frac{\pi}{2}}\frac{\sin^2\theta}{\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{\frac{1}{2}}}\mathrm{d}\theta\\ &= \int_0^{\frac{\pi}{2}}\frac{\alpha\,\sin^2\theta}{\left(\alpha^2\sin^2\theta+\sigma^2\right)^{\frac{1}{2}}}\mathrm{d}\theta,\end{aligned}\tag{A.103}$$

where Equation (i) follows since the density of $|Z|$ is $f_{|Z|}(r)=\int_0^{2\pi}\delta/\pi\exp(-\delta r^2)r\mathrm{d}\theta = 2\delta r\exp(-\delta r^2)$, and Equation (ii) follows from the identity $\int_0^{\infty}r^2\exp(-ar^2)\mathrm{d}r = \sqrt{\pi}/4\cdot a^{-3/2}$, and Equation (iii) is derived in Equation (A.106).

We next derive $\psi_2(\alpha,\sigma^2;\delta)$. From Equation (A.101), we have

$$\begin{aligned}\psi_2(\alpha,\sigma^2;\delta) &= 4\mathbb{E}\left[\left(|Z|-|P|\right)^2\right]\\ &= 4\left(\frac{1+\alpha^2+\sigma^2}{\delta}-2\cdot\mathbb{E}\left\{|ZP|\right\}\right),\end{aligned}$$

where the last step is from $Z\sim\mathcal{CN}(0,1/\delta)$ and $P\sim\mathcal{CN}(0,(\alpha^2+\sigma^2)/\delta)$. We next calculate $\mathbb{E}[|ZP|]$. Again, conditioned on $|Z|$, P is distributed as $P\sim\mathcal{CN}(\alpha|Z|,\sigma^2/\delta)$.

We first calculate $\mathbb{E}[|P|\big||Z|]$:

$$\begin{aligned}
\mathbb{E}\left[|P|\big||Z|\right] &= \int_{\mathbb{C}} |P| \frac{1}{\pi\sigma^2/\delta} \exp\left(-\frac{\left|P-\alpha|Z|\right|^2}{\sigma^2/\delta}\right) \mathrm{d}P \\
&= \int_0^{2\pi}\int_0^{\infty} r \frac{1}{\pi\sigma^2/\delta} \exp\left(-\frac{|re^{\mathrm{i}\theta}-\alpha|Z||^2}{\sigma^2/\delta}\right) \cdot r\mathrm{d}r\mathrm{d}\theta \\
&= \frac{1}{\pi\sigma^2/\delta}\int_0^{2\pi} \exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta \int_0^{\infty} r^2 \exp\left(-\frac{(r-\alpha|Z|\cos\theta)^2}{\sigma^2/\delta}\right)\mathrm{d}r \\
&= \frac{2}{\sqrt{\pi\sigma^2/\delta}}\int_0^{\frac{\pi}{2}} \left(\alpha^2|Z|^2\cos^2\theta + \frac{\sigma^2}{2\delta}\right) \exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta,
\end{aligned} \tag{A.104}$$

where in the last step we used the following indentity

$$\begin{aligned}
\int_0^{\infty} r^2 \exp\left(-\frac{(r-m)^2}{v}\right)\mathrm{d}r &= \frac{mv}{2}\exp\left(-\frac{m^2}{v}\right) \\
&\quad + \sqrt{v\pi}\left(m^2+\frac{v}{2}\right)\Phi\left(\sqrt{\frac{2}{v}}\cdot m\right), \quad \forall m\in\mathbb{R}, v\in\mathbb{R}_+
\end{aligned}$$

and some manipulations similar to those in Equation (A.95). Following the same procedure as that in Equation (A.103), we further calculate $\mathbb{E}[|ZP|]$ as:

$$\begin{aligned}
\mathbb{E}[|ZP|] &= \int_0^{\infty} r\cdot 2r\delta\exp\left(-\delta r^2\right) \\
&\quad \cdot\left(\frac{2}{\sqrt{\pi\sigma^2/\delta}}\int_0^{\frac{\pi}{2}}\left(\alpha^2r^2\cos^2\theta+\frac{\sigma^2}{2\delta}\right)\exp\left(-\frac{\alpha^2r^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta\right)\mathrm{d}r \\
&= \int_0^{\frac{\pi}{2}}\int_0^{\infty}\frac{4\delta^{3/2}}{\sqrt{\pi}\sigma}\left(\alpha^2\cos^2\theta\cdot r^4+\frac{\sigma^2}{2\delta}\cdot r^2\right)\exp\left(-\delta\left(1+\frac{\alpha^2\sin^2\theta}{\sigma^2}\right)r^2\right)\mathrm{d}r\mathrm{d}\theta \\
&= \frac{3\alpha^2}{2\sigma\delta}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{5}{2}}\mathrm{d}\theta + \frac{\sigma}{2\delta}\int_0^{\frac{\pi}{2}}\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{3}{2}}\mathrm{d}\theta,
\end{aligned} \tag{A.105}$$

where in the last step we used the following identities: $\int_0^{\infty} r^4\exp(-ar^2)\mathrm{d}r = 3\sqrt{\pi}/8\cdot a^{-5/2}$ and $\int_0^{\infty} r^2\exp(-ar^2)\mathrm{d}r = \sqrt{\pi}/4\cdot a^{-3/2}$. Finally, using Equation (A.105) we

have

$$\begin{aligned}
\psi_2(\alpha,\sigma^2;\delta) &= 4\left(\frac{1+\alpha^2+\sigma^2}{\delta} - 2\cdot\mathbb{E}\left\{|Z||P|\right\}\right) \\
&\overset{\text{(i)}}{=} 4\left\{\frac{1+\alpha^2+\sigma^2}{\delta} - 2\left[\frac{3\alpha^2}{2\sigma\delta}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{5}{2}}\mathrm{d}\theta\right.\right. \\
&\quad \left.\left. + \frac{\sigma}{2\delta}\int_0^{\frac{\pi}{2}}\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{3}{2}}\mathrm{d}\theta\right]\right\} \\
&= \frac{4}{\delta}\left\{1+\alpha^2+\sigma^2 - \frac{\sigma}{2}\left[\frac{3\alpha^2}{\sigma^2}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{5}{2}}\mathrm{d}\theta\right.\right. \\
&\quad \left.\left. + \int_0^{\frac{\pi}{2}}\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{3}{2}}\mathrm{d}\theta\right]\right\} \\
&\overset{\text{(ii)}}{=} \frac{4}{\delta}\left(1+\alpha^2+\sigma^2 - \sigma\int_0^{\frac{\pi}{2}}\frac{1+2\frac{\alpha^2}{\sigma^2}\sin^2\theta}{\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{\frac{1}{2}}}\mathrm{d}\theta\right) \\
&= \frac{4}{\delta}\left(1+\alpha^2+\sigma^2 - \int_0^{\frac{\pi}{2}}\frac{2\alpha^2\sin^2\theta+\sigma^2}{\left(\alpha^2\sin^2\theta+\sigma^2\right)^{\frac{1}{2}}}\mathrm{d}\theta\right),
\end{aligned}$$

where Equation (i) is from Equation (A.105), and the derivations of Equation (ii) is more involved and are given in Lemma 3.3.

A.3.2 Proof of Theorem 3.1

Since the proof of the real-valued and complex valued signals look similar, for the sake of notational simplicity we present the proof for the real-valued signals. First note that according to [Mondelli and Montanari, 2017, Lemma 13][1] for the smoothed AMP.A algorithm we know that almost surely

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(x_{\epsilon_j,i}^{t+1}(n) - \mathrm{sign}(\alpha_t)\cdot x_i^{\star}\right)^2 = \mathbb{E}(X_{\epsilon_j}^{t+1} - \mathrm{sign}(\alpha_t)\cdot X^{\star})^2,$$

[1]The proof for a more general result was first presented in Javanmard and Montanari [2013]. However, we found Mondelli and Montanari [2017] easier to follow. The reader may also find [Rangan, 2011, Claim 1] and related discussions useful, although no formal proof was provided.

where $X_\epsilon^t = \alpha_{\epsilon,t} X^\star + \sigma_{\epsilon,t} H$ and $X^\star \sim p_X$ is independent of $H \sim \mathcal{N}(0,1)$, and $\alpha_{\epsilon,t}$ and $\sigma_{\epsilon,t}$ satisfy the following iterations:

$$\begin{aligned} \alpha_{\epsilon,t+1} &= \mathbb{E}\left[\partial_z g_\epsilon(P^t, Y)\right], \\ \sigma^2_{\epsilon,t+1} &= \mathbb{E}[g^2_\epsilon(P^t, Y)], \end{aligned}$$

where $Y = |Z| + W$, $P^t = \alpha_{\epsilon,t} Z + \sigma_{\epsilon,t} B$, where $B \sim \mathcal{N}(0, 1/\delta)$ is independent of $Z \sim \mathcal{N}(0, 1/\delta)$ and $W \sim \mathcal{N}(0, 1/\delta)$. It is also straightforward to use an induction step similar to the one presented in the proof of Theorem 1 of Zheng *et al.* [2017] and show that $(\alpha_{\epsilon,t}, \sigma^2_{\epsilon,t}) \to (\alpha_t, \sigma^2_t)$ as $i \to \infty$, where (α_t, σ^2_t) satisfy

$$\begin{aligned} \alpha_{t+1} &= \mathbb{E}\left[\partial_z g(P^t, Y)\right], \\ \sigma^2_{t+1} &= \mathbb{E}[g^2(P^t, Y)]. \end{aligned}$$

A.3.3 Proof of Lemma 3.3

We will only prove Equation (3.14b). Equation (3.14a) can be proved in the same way. The idea is to express the integrals using elliptic integrals defined in Equation (3.11), and then apply known properties of elliptic integrals (Lemma 3.2) to simplify the results. The same tricks in proving Equation (3.14b) are used to derive other related integrals in this chapter. Below, we will provide the full details for the proof of Equation (3.14b), and will not repeat such calculations elsewhere. The LHS of Equation (3.14b) can be rewritten as:

$$\int_0^{\frac{\pi}{2}} \frac{3m}{(1 + m\sin^2\theta)^{\frac{5}{2}}} \mathrm{d}\theta - \int_0^{\frac{\pi}{2}} \frac{3m\sin^2\theta}{(1 + m\sin^2\theta)^{\frac{5}{2}}} \mathrm{d}\theta + \int_0^{\frac{\pi}{2}} \frac{1}{(1 + m\sin^2\theta)^{\frac{3}{2}}} \mathrm{d}\theta \quad \text{(A.106)}$$

The equality in Equation (A.106) can be proved by combining the following identities together with straightfroward manipulations:

$$\text{(i):} \quad \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta = \frac{(m+1)E\left(\frac{m}{1+m}\right) - K\left(\frac{m}{1+m}\right)}{m\sqrt{1+m}}, \tag{A.107a}$$

$$\text{(ii):} \quad \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta = \frac{K\left(\frac{m}{1+m}\right) - E\left(\frac{m}{1+m}\right)}{m\sqrt{1+m}}, \tag{A.107b}$$

$$\text{(iii):} \quad \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta = \frac{1}{\sqrt{1+m}}E\left(\frac{m}{1+m}\right), \tag{A.107c}$$

$$\text{(iv):} \quad \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta = \frac{-(1-m)E\left(\frac{m}{1+m}\right) + K\left(\frac{m}{1+m}\right)}{3m(1+m)^{\frac{3}{2}}}, \tag{A.107d}$$

$$\text{(v):} \quad \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta = \frac{2(m+2)E\left(\frac{m}{1+m}\right) - K\left(\frac{m}{1+m}\right)}{3(1+m)^{\frac{3}{2}}}, \tag{A.107e}$$

where $K(m)$ and $E(m)$ denote the complete elliptic integrals of the first and second kinds (see Equation (3.11)). First, consider the identity (i) in Equation (A.107):

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta &= \frac{1}{m}\int_0^{\frac{\pi}{2}} (1+m\sin^2\theta)^{\frac{1}{2}}\mathrm{d}\theta - \frac{1}{m}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta \\
&\overset{(a)}{=} \frac{1}{m}\left[E(-m) - K(-m)\right] \\
&\overset{(b)}{=} \frac{1}{m}\left[\sqrt{1+m}E\left(\frac{m}{1+m}\right) - \frac{1}{\sqrt{1+m}}K\left(\frac{m}{1+m}\right)\right],
\end{aligned}$$

where (a) is from the definition of $K(m)$ and $E(m)$ in Equation (3.11), and (b) is from Lemma 3.2 (iii).

Identity (ii) can be proved as follows:

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta &= -2\frac{\mathrm{d}}{\mathrm{d}m}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta \\
&= -2\frac{\mathrm{d}}{\mathrm{d}m}K(-m) \\
&\overset{(a)}{=} \frac{(1+m)K(-m) - E(-m)}{m(1+m)} \\
&\overset{(b)}{=} \frac{K\left(\frac{m}{1+m}\right) - E\left(\frac{m}{1+m}\right)}{m\sqrt{1+m}},
\end{aligned} \tag{A.108}$$

where (a) is due to Lemma 3.2 (iv) and (b) is from Lemma 3.2 (iii).

For identity (iii), we have

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta &= \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta - m\cdot\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta \\ &\overset{(a)}{=} K(-m) - m\cdot\frac{(1+m)K(-m)-E(-m)}{m(1+m)} \\ &= \frac{E(-m)}{1+m} \\ &\overset{(b)}{=} \frac{1}{\sqrt{1+m}}E\left(\frac{m}{1+m}\right),\end{aligned} \tag{A.109}$$

where step (a) follows from the third step of Equation (A.108), and step (b) follows from Lemma 3.2 (iii).

Identity (iv) can be proved in a similar way:

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta &= -\frac{2}{3}\cdot\frac{\mathrm{d}}{\mathrm{d}m}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta \\ &\overset{(a)}{=} -\frac{2}{3}\cdot\frac{\mathrm{d}}{\mathrm{d}m}\frac{E(-m)}{1+m} \\ &\overset{(b)}{=} \frac{(1+m)K(-m)-(1-m)E(-m)}{3m(1+m)^2} \\ &\overset{(c)}{=} \frac{-(1-m)E\left(\frac{m}{1+m}\right)-K\left(\frac{m}{1+m}\right)}{3m(1+m)^{\frac{3}{2}}},\end{aligned}$$

where (a) is from the third step of Equation (A.109), step (b) is from Lemma 3.2 (iv) and (c) is from Lemma 3.2 (iii).

Lastly, identity (v) can be proved as follows:

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta &= \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta - m\cdot\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta \\ &\overset{(a)}{=} \frac{E(-m)}{1+m} - m\cdot\frac{(1+m)K(-m)-(1-m)E(-m)}{3m(1+m)^2} \\ &\overset{(b)}{=} \frac{2(m+2)E\left(\frac{m}{1+m}\right)-K\left(\frac{m}{1+m}\right)}{3(1+m)^{\frac{3}{2}}},\end{aligned}$$

where step (a) follows from the derivations of the previous two identities and (b) is again due to Lemma 3.2 (iii).

A.3.4 Proofs omitted in Section 3.3

A.3.4.1 Proof of Lemma 3.9

Part (i): From Equation (3.6), it is easy to verify that $\psi_1(\alpha, \sigma^2)$ is an increasing function of $\alpha > 0$. We now prove its concavity. To this end, we calculate its first and second partial derivatives:

$$\frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha} = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta \cdot \sigma^2}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta, \tag{A.110a}$$

$$\frac{\partial_1^2 \psi_1(\alpha, \sigma^2)}{\partial \alpha^2} = \int_0^{\frac{\pi}{2}} \frac{-3\sin^4\theta \cdot \sigma^2 \alpha}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{5}{2}}} \mathrm{d}\theta < 0, \quad \forall \alpha > 0, \sigma^2 > 0. \tag{A.110b}$$

Hence, $\psi_2(\alpha, \sigma^2)$ is a concave function of α for $\alpha > 0$.

Part (ii): Positivity of ψ_1 is obvious. Also, note that

$$\psi_1(\alpha, \sigma^2) = \int_0^{\pi/2} \frac{\sin^2\theta}{(\sin^2(\theta) + \frac{\sigma^2}{\alpha^2})^{\frac{1}{2}}} d\theta \leq \int_0^{\pi/2} \sin\theta d\theta = 1.$$

Proof of (iii): The claim is a consequence of the concavity of ψ_1 (with respect to α) and the following condition:

$$\left.\frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha}\right|_{\alpha=0} = 1 \Longleftrightarrow \sigma^2 = \frac{\pi^2}{16}.$$

The detailed proof is as follows. First, it is straightforward to verify that $\alpha = 0$ is always a solution to $\alpha = \psi_1(\alpha, \sigma^2)$. Define

$$\Psi_1(\alpha, \sigma^2) \stackrel{\Delta}{=} \psi_1(\alpha, \sigma^2) - \alpha.$$

Since $\Psi_1(\alpha, \sigma^2)$ is a concave function of α (as $\psi_1(\alpha, \sigma^2)$ is concave), $\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}$ is decreasing. Let's first consider $\sigma^2 > \pi^2/16$. In this case we know that

$$\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha} \leq \left.\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}\right|_{\alpha=0} = \left.\frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha}\right|_{\alpha=0} - 1 = \frac{\pi}{4\sigma} - 1 < 0, \tag{A.111}$$

where the second equality can be calculated from Equation (A.110a). Since $\Psi_1(\alpha, \sigma^2)$ is a decreasing function of α and is equal to zero at zero, and it does not have any other solution. Now, consider case $\sigma^2 < \pi^2/16$. It is straightforward to confirm that

$$\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha=0} = \frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha=0} - 1 = \frac{\pi}{4\sigma} - 1 > 0.$$

Furthermore, from Equation (A.110a) we have $\frac{\partial \psi_1(\alpha,\sigma^2)}{\partial \alpha}\Big|_{\alpha\to\infty} = 0$, and so

$$\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha\to\infty} \to -1.$$

Hence, $\Psi_1(\alpha, \sigma^2) = 0$ has exactly one more solution for $\alpha > 0$. Note that since from part (ii) $\psi_1(\alpha, \sigma^2) < 1$, the solution of $\alpha = \psi_1(\alpha, \sigma^2)$ also satisfies $\alpha \leq 1$.

Finally, the strong global attractiveness follows from the fact that ψ_1 is a strictly increasing function of α.

A.3.4.2 Proof of Lemma 3.10

First note that the partial derivative of ψ_2 w.r.t. σ^2 is given by

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta}\left(1 - \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta\right). \tag{A.112}$$

Part (i): Before we proceed, we first comment on the discontinuity of the partial derivative $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}$ at $\sigma^2 = 0$. Note that the formula in Equation (A.112) was derived for non-zero values of σ^2. Naively, one may plug in $\sigma^2 = 0$ in the equation and assume that $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}\Big|_{\alpha=1,\sigma^2=0} = \frac{4}{\delta}$. This is not the case since the integral $\int_0^{\pi/2} \frac{d\theta}{\sin\theta}$ is divergent. It turns out that the derivative $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}$ is a continuous function of σ^2. The technical details can be found in Appendix A.3.6.

Since $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}$ is continuous at $\sigma^2 = 0$, we have

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}\Big|_{\alpha=1,\sigma^2=0} = \lim_{\sigma^2\to 0} \frac{\partial \psi_2(1, \sigma^2; \delta)}{\partial \sigma^2}.$$

Note that if we set $m = 1/\sigma^2$, then from Equation (A.107) we have

$$\frac{\partial \psi_2(1, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta}\left(1 - \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta\right) = \frac{4}{\delta}\left(1 - \frac{1}{2}\sqrt{\frac{m}{1+m}} E\left(\frac{m}{m+1}\right)\right).$$

It is then straightforward to use Lemma 3.2 to prove that

$$\lim_{m\to\infty} \frac{4}{\delta}\left(1-\frac{1}{2}\sqrt{\frac{m}{1+m}}E\left(\frac{m}{m+1}\right)\right) = \frac{2}{\delta}.$$

Hence, $\frac{\partial\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2}\Big|_{\alpha=1,\sigma^2=0} > 1$ for $\delta < 2$.

Part (ii): We first prove that the following equation has at least one solution for any $\alpha \in [0,1]$ and $\delta > 2$:

$$\sigma^2 = \psi_2(\alpha,\sigma^2;\delta), \quad \sigma^2 \in [0,1].$$

It is straightforward to verify that

$$\psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=0} = \frac{4}{\delta}(1-\alpha)^2 \geq 0. \tag{A.113}$$

We next prove our claim by proving the following:

$$\psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=1} < 1, \quad \forall\alpha \in [0,1] \text{ and } \delta > 2. \tag{A.114}$$

From Equation (3.6b), we have

$$\psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=1} < 1 \Longleftrightarrow \underbrace{\int_0^{\frac{\pi}{2}} \frac{2\alpha^2\sin^2\theta+1}{(\alpha^2\sin^2\theta+1)^{\frac{1}{2}}}\mathrm{d}\theta - \alpha^2}_{g(\alpha^2)} > 2-\frac{\delta}{4}. \tag{A.115}$$

We next show that $g(\alpha^2)$ in Equation (A.115) is a concave function of α^2, and hence the minimum can only happen at either $\alpha = 0$ or $\alpha = 1$. The first two derivatives w.r.t. α^2 are given by:

$$\frac{\mathrm{d}g(\alpha^2)}{\mathrm{d}\alpha^2} = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta\left(\alpha^2\sin^2\theta+\frac{3}{2}\right)}{(\alpha^2\sin^2\theta+1)^{\frac{3}{2}}}\mathrm{d}\theta - 1,$$

and

$$\frac{\mathrm{d}^2g(\alpha^2)}{\mathrm{d}(\alpha^2)^2} = -\int_0^{\frac{\pi}{2}} \frac{\sin^4\theta\left(\frac{1}{2}\alpha^2\sin^2\theta+\frac{5}{4}\right)}{(\alpha^2\sin^2\theta+1)^{\frac{5}{2}}}\mathrm{d}\theta < 0.$$

The concavity of $g(\alpha^2)$ implies that its minimum happens at either $\alpha = 0$ or $\alpha = 1$. Hence, to prove Equation (A.115), it suffices to prove that

$$g(0) = \frac{\pi}{2} > 2-\frac{\delta}{4} \quad \text{and} \quad g(1) \approx 1.509 > 2-\frac{\delta}{4},$$

which holds for $\delta > 2$. Hence, Equation (A.115) holds. By combining Equation (A.113) and Equation (A.114) we conclude that $\psi_2(\alpha, \sigma^2; \delta)$ has at least one fixed point between $\sigma^2 = 0$ and $\sigma^2 = 1$. The next step is to prove the uniqueness of this fixed point. For the rest of the proof, we discuss two cases separately: a) $\delta > 4$ and b) $2 < \delta \leq 4$.

(a) $\delta > 4$. Define

$$\Psi_2(\alpha, \sigma^2; \delta) \triangleq \psi_2(\alpha, \sigma^2; \delta) - \sigma^2. \tag{A.116}$$

From Equation (A.112), if $\delta > 4$, then $\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} < 1$, $\forall \sigma^2 > 0$. This means that $\Psi_2(\alpha, \sigma^2; \delta)$ defined in Equation (A.116) is monotonically decreasing in $\sigma^2 > 0$. Hence, the solution to $\Psi_2(\alpha, \sigma^2; \delta) = 0$ is unique. Furthermore, the following property is a direct consequence of the monotonicity of $\Psi_2(\alpha, \sigma^2; \delta)$:

$$\Psi_2(\alpha, \sigma^2; \delta) < 0, \quad \forall 0 < \sigma^2 < F_2(\alpha), \tag{A.117a}$$

and

$$\Psi_2(\alpha, \sigma^2; \delta) > 0 > \sigma^2, \quad \forall F_2(\alpha) < \sigma^2 < 1, \tag{A.117b}$$

where $F_2(\alpha)$ denotes the solution to $\Psi_2(\alpha, \sigma^2; \delta) = 0$.

(b) $2 < \delta \leq 4$. In this case, we will prove that there exists a threshold on σ^2, denoted as $\sigma_\star^2(\alpha; \delta)$ below, such that the following hold:

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} < 1, \ \forall \sigma^2 < \sigma_\star^2(\alpha; \delta) \quad \text{and} \quad \frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} > 1, \ \forall \sigma^2 \in (\sigma_\star^2(\alpha; \delta), \infty). \tag{A.118}$$

This means that $\Psi_2(\alpha, \sigma^2; \delta) = \psi_2(\alpha, \sigma^2; \delta) - \sigma^2$ is strictly decreasing on $\sigma^2 \in (0, \sigma_\star^2(\alpha; \delta))$ and increasing on $\sigma^2 \in (\sigma_\star^2(\alpha; \delta), \infty)$. Note that since we have proved that $\Psi_2(\alpha, \sigma^2; \delta) = 0$ has at least one solution, we conclude that there exist exactly two solutions to $\Psi_2(\alpha, \sigma^2; \delta) = 0$, one in $(0, \sigma_\star^2(\alpha; \delta))$ and the second in $(\sigma_\star^2(\alpha; \delta), \infty)$, if $\Psi_2(\alpha, \sigma^2; \delta)|_{\sigma^2 = \sigma_\star^2(\alpha; \delta)} < 0$. This is the case since $\Psi_2(\alpha, \sigma^2; \delta)|_{\sigma^2 = 1} < 0$ (see Equation (A.114)), and that $\Psi_2(\alpha, \sigma^2; \delta)|_{\sigma^2 = 1} < \Psi_2(\alpha, \sigma^2; \delta)|_{\sigma^2 = \sigma_\star^2(\alpha; \delta)}$ (since the latter is the global minimum of $\Psi_2(\alpha, \sigma^2; \delta)$ in $\sigma^2 \in (0, \infty)$).

Also, it is easy to prove Equation (A.117). In fact, the following holds:

$$\Psi_2(\alpha, \sigma^2; \delta) < 0, \quad \forall 0 < \sigma^2 < F_2(\alpha),$$

and

$$\Psi_2(\alpha, \sigma^2; \delta) > 0 > \sigma^2, \quad \forall F_2(\alpha) < \sigma^2 < \hat{F}_2(\alpha; \delta),$$

where $\hat{F}_2(\alpha; \delta) > 1$ denotes the larger solution to $\Psi_2(\alpha, \sigma^2; \delta) = 0$. See Fig. A.1 for an illustration.

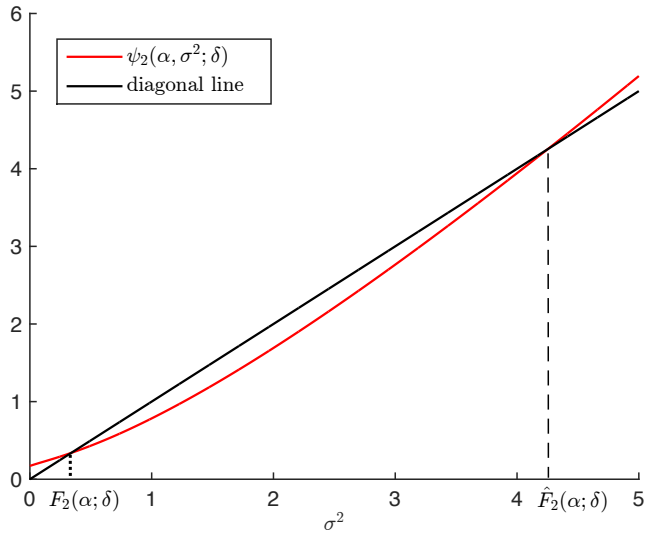

Figure A.1: Plot of $\psi_2(\alpha, \sigma^2; \delta)$ for $\alpha = 0.7$ and $\delta = 2.1$.

From the above discussions, it remains to prove Equation (A.118). To this end, it is more convenient to express Equation (A.112) using elliptic integrals discussed in Section 3.1.4:

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta}\left(1 - \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta\right) \tag{A.119a}$$

$$= \frac{4}{\delta\alpha}\left(\alpha - \underbrace{\frac{1}{2\sqrt{1+s^2}} E\left(\frac{1}{1+s^2}\right)}_{f(s)}\right), \tag{A.119b}$$

where we introduced a new variable $s \triangleq \frac{\sigma}{\alpha}$ and the last step is derived using the identities in Lemma 3.3. Based on Equation (A.119) we can now rewrite Equation (A.118) as

$$\begin{aligned} f(s) &> \alpha\left(1 - \frac{\delta}{4}\right), \quad \forall s < \frac{\sigma_\star(\alpha; \delta)}{\alpha} \\ f(s) &< \alpha\left(1 - \frac{\delta}{4}\right), \quad \forall s \in \left(\frac{\sigma_\star(\alpha; \delta)}{\alpha}, \infty\right). \end{aligned} \tag{A.120}$$

To prove this, we first show that there exists s_* such that $f(s)$ is strictly increasing on $(0, s_*)$ and decreasing on (s_*, ∞), namely,

$$f'(s) > 0, \text{ for } s < s_*, \quad \text{and} \quad f'(s) < 0, \text{ for } s > s_*. \tag{A.121a}$$

s_* is in fact the unique solution to the following equation:

$$2E\left(\frac{1}{1+s_*^2}\right) = K\left(\frac{1}{1+s_*^2}\right). \tag{A.121b}$$

This can be seen from $f'(s)$ derived below:

$$\begin{aligned} f'(s) &= \frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{2\sqrt{1+s^2}}E\left(\frac{1}{1+s^2}\right) \\ &= \frac{s}{2(1+s^2)^{\frac{3}{2}}}\left[K\left(\frac{1}{1+s^2}\right) - 2E\left(\frac{1}{1+s^2}\right)\right]. \end{aligned}$$

Further noting that $E(\cdot)$ is strictly decreasing in $(0,1)$ while $K(\cdot)$ is increasing, we proved Equation (A.121).

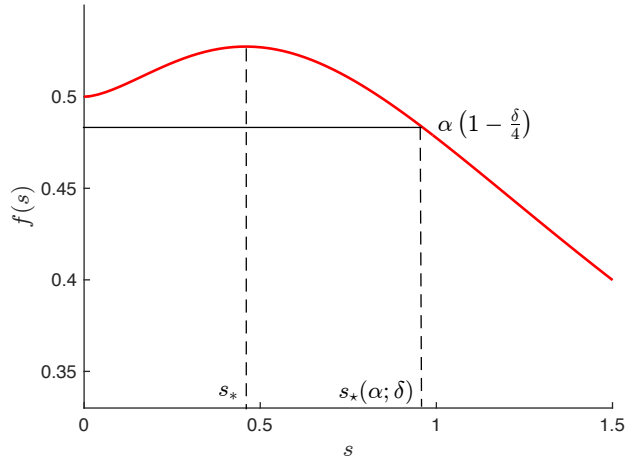

Figure A.2: Illustration of $f(s)$.

Based on the above discussions, we can finally turn to the proof of Equation (A.120). From Equation (A.119b), it is straightforward to verify that $f(0) = \frac{1}{2}$. Therefore, when $\delta > 2$, we have

$$\alpha\left(1-\frac{\delta}{4}\right) \le 1 - \frac{\delta}{4} < \frac{1}{2} = f(0), \quad \forall \delta > 2 \text{ and } 0 \le \alpha \le 1.$$

Hence, the following equation admits a unique solution (denoted as $s_\star(\alpha;\delta)$ below):

$$f(s) = \alpha\left(1-\frac{\delta}{4}\right), \quad \forall \delta > 2 \text{ and } 0 \le \alpha \le 1.$$

See Fig. A.2 for an illustration. Also, from our above discussions on the monotonicity of $f(s)$ it is straightforward to show that

$$f(s) > \alpha\left(1-\frac{\delta}{4}\right),\ \forall s < s_\star(\alpha;\delta) \quad \text{and} \quad f(s) < \alpha\left(1-\frac{\delta}{4}\right),\ \forall s \in (s_\star(\alpha;\delta),\infty),$$

which proves Equation (A.120) by setting $\sigma_\star(\alpha;\delta) \stackrel{\Delta}{=} \alpha \cdot s_\star(\alpha;\delta)$. This proves Equation (A.118), which completes the proof.

Part (iii): We will prove a stronger result: $\psi_2 \le 4/\delta$. From Equation (3.6b), $\psi_2(\alpha,\sigma^2;\delta) \le 4/\delta$ is equivalent to

$$\alpha^2 + \sigma^2 - \int_0^{\frac{\pi}{2}} \frac{2\alpha^2\sin^2\theta + \sigma^2}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta \le 0,$$

which can be further reformulated as

$$\alpha^2 \le \int_0^{\frac{\pi}{2}} \frac{2\alpha^2\sin^2\theta}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta + \sigma^2\left(\int_0^{\frac{\pi}{2}} \frac{1}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta - 1\right). \tag{A.122}$$

For $0 \le \alpha \le 1$ and $\sigma^2 \le \sigma^2_{\max}$ we have

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \frac{1}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta &\ge \int_0^{\frac{\pi}{2}} \frac{1}{\left(\sin^2\theta+\sigma^2_{\max}\right)^{\frac{1}{2}}}\mathrm{d}\theta, \\ &\stackrel{(a)}{=} \int_0^{\frac{\pi}{2}} \frac{1}{\left(\sin^2\theta+\frac{4}{\delta_{\mathrm{AMP}}}\right)^{\frac{1}{2}}}\mathrm{d}\theta \\ &\approx 1.09 > 1,\end{aligned} \tag{A.123}$$

where step (a) from $\sigma^2_{\max} = \max\{1, 4/\delta\} \ge \max\{1, 4/\delta_{\mathrm{AMP}}\} = 4/\delta_{\mathrm{AMP}} \approx 1.6$. Due to Equation (A.123), to prove Equation (A.122), it suffices to prove

$$\alpha^2 \le \int_0^{\frac{\pi}{2}} \frac{2\alpha^2\sin^2\theta}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta,$$

or

$$1 \leq \int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta,$$

which, similar to Equation (A.123), can be proved by

$$\int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta \geq \int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta}{\left(\sin^2\theta+\frac{4}{\delta_{\mathrm{AMP}}}\right)^{\frac{1}{2}}}\mathrm{d}\theta \approx 1.02 > 1.$$

Part (iv): We bound the partial derivative of $\psi_2(\alpha,\sigma^2;\delta)$ for $\sigma^2 \in [0,\sigma^2_{\max}]$ as:

$$\begin{aligned}
\frac{\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2} &= \frac{4}{\delta}\left(1-\frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{3}{2}}}\mathrm{d}\theta\right)\\
&\overset{(a)}{\leq} \frac{4}{\delta}\left(1-\frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\theta^2+\sigma^2)^{\frac{3}{2}}}\mathrm{d}\theta\right)\\
&\overset{(b)}{=} \frac{4}{\delta}\left(1-\frac{1}{2}\int_0^{\frac{\pi}{2\sigma}} \frac{1}{(\tilde{\theta}^2+1)^{\frac{3}{2}}}\mathrm{d}\tilde{\theta}\right)\\
&\overset{(c)}{\leq} \frac{4}{\delta_{\mathrm{AMP}}}\left(1-\frac{1}{2}\int_0^{\frac{\pi}{2\sqrt{\frac{4}{\delta_{\mathrm{AMP}}}}}} \frac{1}{(\tilde{\theta}^2+1)^{\frac{3}{2}}}\mathrm{d}\tilde{\theta}\right)\\
&\approx 0.98 < 1,
\end{aligned} \tag{A.124}$$

where step (a) follows from the constraint $0 \leq \alpha \leq 1$ and the inequality $\sin\theta \leq \theta$; (b) is due to the variable change $\tilde{\theta} = \theta/\sigma$; (c) is a consequence of the constraint $\sigma^2 \leq \sigma^2_{\max} = \max\{1, 4/\delta\} \leq \max\{1, 4/\delta_{\mathrm{AMP}}\} = 4/\delta_{\mathrm{AMP}}$. As a result of Equation (A.124), $\Psi_2(\alpha,\sigma^2;\delta) = \psi_2(\alpha,\sigma^2;\delta) - \sigma^2$ is decreasing in $\sigma^2 \in [0,\sigma^2_{\max}]$. It is easy to verify that $\psi_2(0,\alpha;\delta) \geq 0$ for $\alpha \in [0,1]$. Further, Lemma 3.10 (iii) implies that

$$\psi_2(\sigma^2_{\max},\alpha;\delta) - \sigma^2_{\max} \leq 0.$$

Hence, there exists a unique solution (which we denote as $F_2(\alpha)$) to the following equation:

$$\psi_2(\sigma,\alpha;\delta) = \sigma^2, \quad 0 \leq \sigma^2 \leq \sigma^2_{\max}.$$

Finally, the property in Equation (3.23) is a direct consequence of the fact that $\Psi_2(\alpha,\sigma^2;\delta) = \psi_2(\alpha,\sigma^2;\delta) - \sigma^2$ is a decreasing function of $\sigma^2 \leq \sigma^2_{\max}$.

Part (v): In Equation (A.119), we have derived the following:

$$\frac{\psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta \alpha} \left(\alpha - f(s) \right),$$

where $s \triangleq \frac{\sigma}{\alpha}$. From Equation (A.119b), we see that $\psi_2(\alpha, \sigma^2; \delta)$ is an increasing function of σ^2 if the following holds:

$$\alpha > f(s).$$

Further, Equation (A.121) implies that the maximum of $f(s)$ happens at s_*, i.e.,

$$\max_{s>0} \; f(s) = \frac{1}{2\sqrt{1+s_*^2}} E\left(\frac{1}{1+s_*^2}\right) \triangleq \alpha_*, \tag{A.125}$$

where s_*^2 is the unique solution to

$$2E\left(\frac{1}{1+s_*^2}\right) = K\left(\frac{1}{1+s_*^2}\right).$$

Clearly, $\alpha > \alpha_*$ immediately implies $\alpha > f(s)$, which further guarantees that $\psi_2(\alpha, \sigma^2; \delta)$ is monotonically increasing on $\sigma^2 > 0$. Finally, the strong global attractiveness of $F_2(\alpha)$ is a direct consequence of part (iv) of this lemma together with the monotonicity of ψ_2.

A.3.4.3 Proof of Lemma 3.11

Part (i): We first verify $F_1(0) = 1$ and $\lim_{\sigma^2 \to \frac{\pi^2}{16}-} F_1(\sigma^2) = 0$. First, $F_1(0) = 1$ can be seen from the following facts: (a) $\psi_1(\alpha, 0) = 1$ for $\alpha > 0$, see Equation (3.6a); and (b) By definition, $F_1(0)$ is the non-zero solution to $\alpha = \psi_1(\alpha, 0)$. Then, by Lemma 3.9 (iii) and continunity of ψ_1, we know F_1 is continuous on $[0, \frac{\pi^2}{16})$, and further $\lim_{\sigma^2 \to \frac{\pi^2}{16}-} F_1(\sigma^2) = 0$ since $\sigma^2 = \frac{\pi^2}{16}$ corresponds to a case where the non-negative solution to $\psi_1(\alpha, \sigma^2) = \alpha$ decreases to zero. Next, we prove the monotonicity of F_1. Note that

$$F_1(\sigma^2) = \psi_1(F_1(\sigma^2), \sigma^2),$$

Differentiation w.r.t. σ^2 yields

$$F_1'(\sigma^2) = \partial_2 \psi_1(F_1(\sigma^2), \sigma^2) + \partial_1 \psi_1(F_1(\sigma^2), \sigma^2) \cdot F_1'(\sigma^2),$$

where $\partial_2\psi_1(F_1(\sigma^2),\sigma^2) \triangleq \frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\sigma^2}\Big|_{\alpha=F_1(\sigma^2)}$ and $\partial_1\psi_1(F_1(\sigma^2),\sigma^2) \triangleq \frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=F_1(\sigma^2)}$. Hence,

$$\left[1-\partial_1\psi_1(F_1(\sigma^2),\sigma^2)\right]\cdot F_1'(\sigma^2) = \partial_2\psi_1(F_1(\sigma^2),\sigma^2). \tag{A.126}$$

We have proved in Equation (A.111) that $\frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=0} < 1$ when $\sigma^2 < \frac{\pi^2}{16}$. Together with the concavity of ψ_1 w.r.t. α (cf. Lemma 3.9 (i)), we have

$$\frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=F_1(\sigma^2)} < \frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=0} < 1, \quad \forall\sigma^2 < \frac{\pi^2}{16}. \tag{A.127}$$

Further, from Equation (3.6a), it is straightforward to see that ψ_1 is a strictly decreasing function of σ^2, and thus

$$\partial_2\psi_1(F_1(\sigma^2),\sigma^2) = \frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=F_1(\sigma^2)} < 0. \tag{A.128}$$

Substituting Equation (A.127) and Equation (A.128) into Equation (A.126), we obtain

$$F_1'(\sigma^2) < 0, \quad \forall\sigma^2 < \frac{\pi^2}{16}.$$

Proof of (ii): By Lemma 3.10 (ii) and continuity of ψ_2, it is straightforward to check that F_2 is continuous. Moreover, we have proved that $\sigma^2 = F_2(\alpha;\delta)$ is the unique solution to the following equation (for $\delta > 2$):

$$\sigma^2 = \frac{4}{\delta}\left(\alpha^2+\sigma^2+1-\int_0^{\frac{\pi}{2}}\frac{2\alpha^2\sin^2\theta+\sigma^2}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta\right), \quad \sigma^2\in[0,1]. \tag{A.129}$$

When $\alpha = 0$, Equation (A.129) reduces

$$\sigma^2 = \frac{4}{\delta}\left(\sigma^2+1-\frac{\pi}{2}\sigma\right), \quad \sigma^2\in[0,1],$$

which has two possible solutions (for $\delta \neq 4$):

$$\sigma_1 = \frac{-\pi+\sqrt{\pi^2+4(\delta-4)}}{\delta-4} \quad \text{and} \quad \sigma_2 = \frac{-\pi-\sqrt{\pi^2+4(\delta-4)}}{\delta-4}.$$

(For the special case $\delta = 4$, $\sigma_1 = 2/\pi$.) However, σ_2 is invalid due to our constraint $0 < \sigma^2 < 1$. This can be seen as follows. First, $\sigma_2 < 0$ for $\delta > 4$ and hence invalid.

When $2 < \delta < 4$, we have

$$\sigma_2 = \frac{\pi + \sqrt{\pi^2 - 4(4-\delta)}}{4-\delta} > \frac{\pi}{4-\delta} > 1.$$

Hence, $F_2(0;\delta) = \sigma_1$. When $\alpha = 1$, Equation (A.129) becomes:

$$\sigma^2 = \frac{4}{\delta}\left(2 + \sigma^2 - \int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta + \sigma^2}{(\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta\right), \quad \sigma^2 \in [0,1].$$

It is straightforward to verify that $\sigma^2 = 0$ is a solution. Also, from Lemma 3.10 (ii), $\sigma^2 = 0$ is a also the unique solution. Hence, $F_2(1;\delta) = 0$.

A.3.4.4 Proof of Lemma 3.5

In Lemma 3.10, we have proved that $F_2(\alpha;\delta)$ is the unique globally attracting fixed point of ψ_2 in $\sigma^2 \in [0,1]$ (for $\delta > 2$), and from Equation (3.22) we have

$$\sigma^2 > F_2(\alpha;\delta) \Longleftrightarrow \psi_2(\alpha,\sigma^2;\delta) < \sigma^2, \quad \sigma^2 \in [0,1]. \tag{A.130}$$

Here, our objective is to prove that $F_1^{-1}(\alpha) < F_2(\alpha;\delta)$ holds for any $\alpha \in (0,1)$ when $\delta \geq \delta_{\mathrm{AMP}}$. From Equation (A.130) and noting that $F_1^{-1}(\alpha) \leq \pi^2/16 < 1$ (from Lemma 3.11), our problem can be reformulated as proving the following inequality (for $\delta > \delta_{\mathrm{AMP}}$):

$$\psi_2(\alpha, F_1^{-1}(\alpha);\delta) < F_1^{-1}(\alpha), \quad \forall \alpha \in (0,1). \tag{A.131}$$

Since $\psi_2(\alpha, F_1^{-1}(\alpha);\delta)$ is a strictly decreasing function of δ (see Equation (3.6b)), it suffices to prove that Equation (A.131) holds for $\delta = \delta_{\mathrm{AMP}}$:

$$\psi_2(\alpha, F_1^{-1}(\alpha);\delta_{\mathrm{AMP}}) < F_1^{-1}(\alpha), \quad \forall \alpha \in (0,1). \tag{A.132}$$

We now make some variable changes for Equation (A.132). From Equation (3.6a), ψ_1 in can be rewritten as the following for $\alpha > 0$:

$$\psi_1(\alpha,\sigma^2) = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + \frac{\sigma^2}{\alpha^2}\right)^{\frac{1}{2}}} \mathrm{d}\theta.$$

By definition, $F_1(\sigma^2)$ is the solution to $\alpha = \psi_1(\alpha, \sigma^2)$, and hence the following holds:

$$\alpha = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + \frac{F_1^{-1}(\alpha)}{\alpha^2}\right)^{\frac{1}{2}}} \mathrm{d}\theta.$$

At this point, it is more convenient to make the following variable change:

$$s \triangleq \frac{\sqrt{F_1^{-1}(\alpha)}}{\alpha}, \tag{A.133}$$

from which we get

$$\alpha = \phi_1(s) \triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta. \tag{A.134}$$

Notice that $\phi_1 : \mathbb{R}_+ \mapsto [0, 1]$ is a monotonically decreasing function, and it defines a one-to-one map between α and s. From the above definitions, we have

$$F_1^{-1}(\alpha) = s^2\alpha^2 = s^2\phi_1^2(s), \tag{A.135}$$

where the first equality is from Equation (A.133) and the second step from Equation (A.134). Using the relationship in Equation (A.135), we can reformulate the inequality in Equation (A.132) into the following equivalent form:

$$\psi_2(\phi_1(s), s^2\phi_1^2(s); \delta_{\mathrm{AMP}}) < s^2\phi_1^2(s), \quad \forall s > 0. \tag{A.136}$$

Substituting Equation (A.134) and Equation (3.6b) into Equation (A.136) and after some manipulations, we can finally write our objective as:

$$\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{(1 - \gamma s^2)\sin^2\theta + s^2}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta > 1, \quad \forall s > 0. \tag{A.137}$$

where we defined

$$\gamma \triangleq 1 - \frac{\delta_{\mathrm{AMP}}}{4} = 2 - \frac{16}{\pi^2}. \tag{A.138}$$

In the next two subsections, we prove Equation (A.137) for $s^2 > 0.07$ and $s^2 \leq 0.07$ using different techniques.

(i) Case I: We make another variable change:

$$t \triangleq \frac{1}{s^2}.$$

Using the variable t, we can rewrite Equation (A.137) into the following:

$$G(t) \triangleq \frac{g_1(t)}{g_2(t)} - \frac{1}{g_2^2(t)} \geq \gamma, \quad \forall t \in [0, 14.3). \tag{A.139a}$$

where γ is defined in Equation (A.138), and

$$g_1(t) \triangleq \int_0^{\frac{\pi}{2}} (1 + t\sin^2\theta)^{\frac{1}{2}} d\theta, \tag{A.139b}$$

$$g_2(t) \triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1 + t\sin^2\theta)^{\frac{1}{2}}} d\theta. \tag{A.139c}$$

Notice that if we could prove Equation (A.139a) for $t < 14.3$, we would have proved Equation (A.137) for $s^2 > 0.07$, since $14.3 > 1/0.07 \approx 14.28$. For the ease of later discussions, we define

$$g_3(t) \triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^4\theta}{(1 + t\sin^2\theta)^{\frac{3}{2}}} d\theta,$$

$$g_4(t) \triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^6\theta}{(1 + t\sin^2\theta)^{\frac{5}{2}}} d\theta.$$

The following identities related to $\{g_1(t), g_2(t), g_3(t), g_4(t)\}$ will be used in our proof:

$$\begin{aligned} g_1'(t) &= \frac{1}{2} g_2(t), \\ g_2'(t) &= -\frac{1}{2} g_3(t), \\ g_3'(t) &= -\frac{3}{2} g_4(t). \end{aligned} \tag{A.140}$$

We now prove Equation (A.139a). First, it is straightforward to verify that equality holds for Equation (A.139a) at $t = 0$, i.e.,

$$G(0) = \gamma. \tag{A.141}$$

Hence, to prove that $G(t) \geq \gamma$ for $t \in [0, 14.3)$, it is sufficient to prove that $G(t)$ is an increasing function of t on $t \in [0, 14.3)$. To this end, we calculate the

derivative of $G(t)$:

$$\begin{aligned} G'(t) &= \frac{g_1'(t)g_2(t) - g_1(t)g_2'(t)}{g_2^2(t)} - \left(\frac{-2g_2'(t)}{g_2^3(t)}\right) \\ &\overset{(a)}{=} \frac{\frac{1}{2}g_2^2(t) + \frac{1}{2}g_1(t)g_3(t)}{g_2^2(t)} - \frac{g_3(t)}{g_2^3(t)} \\ &= 1 + \frac{1}{2}\frac{g_1(t)g_3(t)}{g_2^2(t)} - \frac{g_3(t)}{g_2^3(t)} \\ &= \frac{1}{2}\frac{g_3(t)}{g_2^3(t)}\Bigg(\underbrace{\frac{g_2^3(t)}{g_3(t)}}_{G_1(t)} + \underbrace{g_1(t)g_2(t)}_{G_2(t)} - 2\Bigg), \end{aligned}$$

where step (a) follows from the identities listed in Equation (A.140). Since $g_3(t) > 0$, we have

$$G'(t) > 0 \Longleftrightarrow G_1(t) + G_2(t) - 2 > 0.$$

It remains to prove that $G_1(t)+G_2(t)-2 > 0$ for $t < 14.3$. Our numerical results suggest that $G_1(t)+G_2(t)$ is a monotonically decreasing function for $t > 0$, and $G_1(t) + G_2(t) \to 2$ as $t \to \infty$. However, directly proving the monotonicity of $G_1(t) + G_2(t)$ seems to be quite complicated. We use a different strategy here. We will prove that (at the end of this section)

- $G_1(t)$ is monotonically increasing;
- $G_2(t)$ is monotonically decreasing.

As a consequence, the following hold true for any $c_2 > c_1 > 0$:

$$G_1(t) + G_2(t) - 2 \geq G_1(c_1) + G_2(c_2) - 2, \quad \forall t \in [c_1, c_2].$$

Hence, if we verify that $G_1(c_1)+G_2(c_2)-2 > 0$, we will be proving the following:

$$G_1(t) + G_2(t) - 2 > 0, \quad \forall t \in [c_1, c_2].$$

To this end, we verify that $G_1(c_1) + G_2(c_2) - 2 > 0$ hold for a sequence of c_1 and c_2: $[c_1, c_2] = [0, 0.49]$, $[c_1, c_2] = [0.49, 1.08]$, $[c_1, c_2] = [1.08, 1.78]$, $[c_1, c_2] = [1.78, 2.56]$, $[c_1, c_2] = [2.56, 3.47]$, $[c_1, c_2] = [3.47, 4.47]$, $[c_1, c_2] = [4.47, 5.56]$, $[c_1, c_2] = [5.56, 6.77]$, $[c_1, c_2] = [6.67, 8.08]$, $[c_1, c_2] = [8.08, 9.5]$, $[c_1, c_2] = [9.5, 11]$,

$[c_1, c_2] = [11, 12.6]$, $[c_1, c_2] = [12.6, 14.3]$. Combining all the above results proves

$$G_1(t) + G_2(t) - 2 > 0, \quad \forall t \in [0, 14.3].$$

From the above discussions, it only remains to prove the monotonicity of $G_1(t)$ and $G_2(t)$. Consider $G_1(t)$ first:

$$\begin{aligned} G_1'(t) &= \left(\frac{g_2^3(t)}{g_3(t)}\right)' \\ &= \frac{3g_2^2(t)g_2'(t)g_3(t) - g_2^3(t)g_3'(t)}{g_3^2(t)} \\ &= \frac{-\frac{3}{2}g_2^2(t)g_3^2(t) + \frac{3}{2}g_2^3(t)g_4(t)}{g_3^2(t)} \\ &= -\frac{3}{2}g_2^2(t) + \frac{3}{2}\frac{g_2^3(t)g_4(t)}{g_3^2(t)} \\ &= \frac{3}{2}\frac{g_2^2(t)}{g_3^2(t)} \cdot [-g_3^2(t) + g_2(t)g_4(t)]. \end{aligned} \tag{A.142}$$

Applying the Cauchy-Schwarz inequality yields:

$$\begin{aligned} g_2(t)g_4(t) &= \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+t\sin^2\theta)^{\frac{1}{2}}}d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{\sin^6\theta}{(1+t\sin^2\theta)^{\frac{5}{2}}}d\theta \\ &\geq \left(\int_0^{\frac{\pi}{2}} \frac{\sin^4\theta}{(1+t\sin^2\theta)^{\frac{3}{2}}}d\theta\right)^2 \\ &= g_3^2(t). \end{aligned} \tag{A.143}$$

Combining Equation (A.142) and Equation (A.143), we proved that $G_1'(t) \geq 0$, and therefore $G_1(t)$ is monotonically increasing. For $G_2(t)$, we have

$$\begin{aligned} G_2'(t) &= g_1'(t)g_2(t) + g_1(t)g_2'(t) \\ &= \frac{1}{2}g_2^2(t) + g_1(t)\left(-\frac{1}{2}g_3(t)\right) \\ &= \frac{1}{2}[g_2^2(t) - g_1(t)g_3(t)]. \end{aligned}$$

Again, using Cauchy-Schwarz we have

$$\begin{aligned} g_1(t)g_3(t) &= \int_0^{\frac{\pi}{2}} (1+t\sin^2\theta)^{\frac{1}{2}} d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{\sin^4\theta}{(1+t\sin^2\theta)^{\frac{3}{2}}} d\theta \\ &\geq \left(\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+t\sin^2\theta)^{\frac{1}{2}}} d\theta \right)^2 \\ &= g_2^2(t). \end{aligned}$$

Combining the previous two equations leads to $G_2'(t) \geq 0$, which completes our proof.

(ii) Case II: We next prove Equation (A.137) for $s^2 \leq 0.07$, which is based on a different strategy. Some manipulations of the RHS of Equation (A.137) yields:

$$\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(\sin^2\theta+s^2)^{\frac{1}{2}}} d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{(1-\gamma s^2)\sin^2\theta + s^2}{(\sin^2\theta+s^2)^{\frac{1}{2}}} d\theta = \frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2}, \tag{A.144a}$$

where $E(\cdot)$, $K(\cdot)$ and $T(\cdot)$ are elliptic integrals defined in Equation (3.11), γ is a constant defined in Equation (A.138), and x is a new variable:

$$x \triangleq \frac{1}{1+s^2}. \tag{A.144b}$$

From our reformulation in Equation (A.144), the inequality in Equation (A.137) for $s^2 < 0.07$ becomes

$$\frac{E(x)T(x)}{x} - \gamma\frac{(1-x)T^2(x)}{x^2} > 1, \quad x \in [0.93, 1). \tag{A.145}$$

Note that $0.93 < 1/(1+0.07)$ and thus proving the above inequality for $x \in [0.93, 1)$ is sufficient to prove the original inequality for $s^2 \leq 0.07$ (note that $x \triangleq 1/(1+s^2)$, see Equation (A.144b)).

With some further calculations, Equation (A.145) can be reformulated as

$$\frac{x}{T^2(x)}\frac{E(x)T(x)-x}{(1-x)} > \gamma, \quad x \in [0.93, 1). \tag{A.146}$$

The following inequality is due to [Anderson and Vamanamurthy, 1985, Eqn. (1)]

$$T(x) < x < 1, \quad \forall x \in (0,1).$$

Hence,

$$\frac{x}{T^2(x)}\frac{E(x)T(x)-x}{(1-x)} > \frac{E(x)T(x)-x}{1-x}, \quad \forall x \in (0,1),$$

and to prove Equation (A.146) it suffices to prove the following

$$\frac{E(x)T(x)-x}{1-x} > \gamma, \quad \forall x \in [0.93,1). \tag{A.147}$$

To this end, we will prove that the LHS of Equation (A.147) is a strictly increasing function of $x \in [0.93,1)$. If this is true, we would have for all $x \in [0.93,1)$

$$\frac{E(x)T(x)-x}{1-x} > \frac{E(x)T(x)-x}{1-x}|_{x=0.93} \approx 0.385 > \gamma = 2 - \frac{16}{\pi^2} \approx 0.3789.$$

We next prove the monotonicity of $\frac{E(x)T(x)-x}{1-x}$. From the identities in Lemma 3.2, we derive the following

$$[E(x)T(x)-x]' = \frac{E^2(x) - 2(1-x)E(x)K(x) + (1-x)K^2(x)}{2x} - 1.$$

Hence, to prove that $\frac{E(x)T(x)-x}{1-x}$ is monotonically increasing, it is sufficient to prove the following inequality:

$$\begin{aligned} 0 \quad < \quad & \left(\frac{E^2(x) - 2(1-x)E(x)K(x) + (1-x)K^2(x)}{2x} - 1\right)(1-x) \\ & -[E(x)T(x)-x](-1). \end{aligned} \tag{A.148}$$

Now, substituting $T(x) = E(x) - (1-x)K(x)$ into Equation (A.148) and after some manipulations, we finally reformulate the inequality to be proved into the following form:

$$T(x)^2 > 2x - xE^2(x).$$

It can be verified that equality holds at $x = 1$. We next prove that $T(x)^2 + xE(x)^2 - 2x$ is monotonically decreasing on $[0.93,1)$. We differentiate once more:

$$(T(x)^2 + xE(x)^2 - 2x)' = 2E(x)^2 - (1-x)K(x)^2 - 2.$$

Our problem boils down to proving $2E(x)^2 - (1-x)K(x)^2 - 2 < 0$ for $x \in [0.93, 1)$. We can verify that $2E(x)^2 - (1-x)K(x)^2 - 2 = 0$ holds at $x = 1$. We finish by showing that $2E(x)^2 - (1-x)K(x)^2 - 2$ is monotonically increasing in $x \in [0.93, 1)$. To this end, we differentiate again:

$$\begin{aligned}[2E(x)^2 - (1-x)K(x)^2 - 2]' &= \frac{K(x)^2 - 3E(x)K(x) + 2E(x)^2}{x} \\ &= \frac{\left[K(x) - \frac{3}{2}E(x)\right]^2 - \frac{1}{2}E(x)^2}{x}.\end{aligned} \tag{A.149}$$

We note that $K(x) - \left(\frac{3}{2} + \frac{1}{\sqrt{2}}\right) E(x)$ is a monotonically increasing function in (0,1) since $K(x)$ is monotonically increasing and $E(x)$ is monotonically decreasing. We verify that $K(x) - \left(\frac{3}{2} + \frac{1}{\sqrt{2}}\right) E(x) > 0$ when $x \geq 0.93$. Hence,

$$K(x) - \left(\frac{3}{2} + \frac{1}{\sqrt{2}}\right) E(x) > 0, \quad \forall x \in [0.93, 1),$$

and therefore

$$\left(K(x) - \frac{3}{2}E(x)\right)^2 > \frac{1}{2}E(x)^2, \quad \forall x \in [0.93, 1). \tag{A.150}$$

Substituting Equation (A.150) into Equation (A.149), we prove that $[2E(x)^2 - (1-x)K(x)^2 - 2]' > 0$ for $x \in [0.93, 1)$, which completes the proof.

A.3.4.5 Proof of Lemma 3.12

Define $\mathcal{X} \triangleq \{(\alpha, \sigma^2) | \alpha > 0, \sigma^2 > 0\}$. Let $\mathcal{Y}$ be the image of $\mathcal{X}$ under the SE map in Equation (3.6). We will prove that the following holds for an arbitrary $C \in [0, 1]$:

$$L(C; \delta) = \min_{(\hat{\alpha}, \hat{\sigma}^2) \in \mathcal{X}} \psi_2(\hat{\alpha}, \hat{\sigma}^2; \delta), \tag{A.151}$$

where $(\hat{\alpha}, \hat{\sigma}^2)$ satisfies the constraint

$$\psi_1(\hat{\alpha}, \hat{\sigma}^2) = C.$$

If Equation (A.151) holds, we would have proved Equation (3.27). To see this, consider arbitrary (α, σ^2) such that $\psi_1(\alpha, \sigma^2) = C$. Then, we have

$$L\left[\psi_1(\alpha,\sigma^2);\delta\right] \overset{(a)}{=} \min_{(\hat{\alpha},\hat{\sigma}^2)} \ \psi_2(\hat{\alpha},\hat{\sigma}^2;\delta) \overset{(b)}{\leq} \psi_2(\alpha,\sigma^2;\delta),$$

where step (a) follows from Equation (A.151) and $\psi_1(\alpha, \sigma^2) = C$, and step (b) holds since the choice $\hat{\alpha} = \alpha$ and $\hat{\sigma}^2 = \sigma^2$ is feasible for the constraint $\psi_1(\hat{\alpha}, \hat{\sigma}^2) = \psi_1(\alpha, \sigma^2)$. This is precisely Equation (3.27).

We now prove Equation (A.151). From Equation (3.6a) we have

$$\psi_1(\alpha,\sigma^2) = \int_0^{\pi/2} \frac{\alpha \sin^2\theta}{(\alpha^2\sin^2\theta + \sigma^2)^{1/2}} d\theta.$$

Furthermore, from the definition of ϕ_1 in Equation (3.26a) we have

$$\psi_1(\hat{\alpha},\hat{\sigma}^2) = \phi_1\left(\frac{\hat{\sigma}}{\hat{\alpha}}\right) = C \Longrightarrow s \triangleq \frac{\hat{\sigma}}{\hat{\alpha}} = \phi_1^{-1}(C). \tag{A.152}$$

Similarly, from Equation (3.6b), i.e. the definition of ψ_2, and the definition of ϕ_2 in Equation (3.26b), we can express $\psi_2(\hat{\alpha}, \hat{\sigma}^2; \delta)$ as

$$\begin{aligned}\psi_2(\hat{\alpha},\hat{\sigma}^2;\delta) &= \frac{4}{\delta}\left[\hat{\alpha}^2 + \hat{\sigma}^2 + 1 - \hat{\alpha}\cdot\phi_2\left(\frac{\hat{\sigma}}{\hat{\alpha}}\right)\right] \\ &= \frac{4}{\delta}\left[(1+s^2)\hat{\alpha}^2 + 1 - \hat{\alpha}\cdot\phi_2(s)\right].\end{aligned}$$

From Equation (A.152), we see that fixing $\psi_1(\hat{\alpha}, \hat{\sigma}^2) = C$ is equivalent to fixing $s = \phi_1^{-1}(C)$. Further, for a fixed s, $\psi_2(\hat{\alpha}, \hat{\sigma}^2)$ is a quadratic function of $\hat{\alpha}$, and the minimum happens at

$$\hat{\alpha}_{\min} = \frac{\phi_2(s)}{2(1+s^2)} = \frac{\phi_2(\phi_1^{-1}(C))}{2\left[1 + \left(\phi_1^{-1}(C)\right)^2\right]},$$

and $\psi_2(\hat{\alpha}_{\min}, \hat{\sigma}^2; \delta)$ is

$$\psi_2(\hat{\alpha}_{\min},\hat{\sigma}^2;\delta) = \frac{4}{\delta}\left(1 - \frac{\phi_2^2(s)}{4(1+s^2)}\right) = \frac{4}{\delta}\left(1 - \frac{\phi_2^2(\phi_1^{-1}(C))}{4\left(1+\left[\phi_1^{-1}(C)\right]^2\right)}\right) = L\left(C;\delta\right),$$

where the last step is from the definition of L is Equation (3.25). This completes the

proof.

A.3.4.6 Proof of Lemma 3.13

Recall from Equation (3.25) that $L(\alpha;\delta)$ is defined as

$$\begin{aligned} L(\alpha;\delta) &\triangleq \frac{4}{\delta}\left(1-\frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4(1+(\phi_1^{-1}(\alpha))^2)}\right) \\ &= \frac{4}{\delta}\left(1-I_2[\phi_1^{-1}(\alpha)]\right), \end{aligned}$$

where $I_2:\mathbb{R}_+\mapsto\mathbb{R}_+$ is defined as

$$I_2(s) \triangleq \frac{\phi_2^2(s)}{4(1+s^2)}. \tag{A.153}$$

From Equation (3.26a), it is easy to see that $\phi_1(s)$ is a decreasing function. Hence, to prove that $L(\alpha;\delta)$ is a decreasing function of α, it suffices to prove that $I_2(s)$ is strictly decreasing.

Substituting Equation (3.26b) into Equation (A.153) yields:

$$\begin{aligned} I_2(s) &= \frac{\phi_2^2(s)}{4(1+s^2)} \\ &= \frac{1}{4(1+s^2)}\left(\int_0^{\frac{\pi}{2}}\frac{2\sin^2\theta+s^2}{\left(\sin^2\theta+s^2\right)^{\frac{1}{2}}}\right)^2 \\ &\overset{(a)}{=} \frac{1}{4}\left[2E\left(\frac{1}{1+s^2}\right)-\frac{s^2}{1+s^2}K\left(\frac{1}{1+s^2}\right)\right]^2 \\ &= \frac{1}{4}\left[2E(x)-(1-x)K(x)\right]^2, \end{aligned}$$

where step (a) is obtained through similar calculations as those in Equation (A.107), and in the last step we defined $x=\frac{1}{1+s^2}$. Hence, to prove that $I_2(s)$ is a decreasing function of s, it suffices to prove that $[2E(x)-(1-x)K(x)]^2$ is an increasing function of x. Further, $2E(x)-(1-x)K(x)=T(x)+E(x)>0$ (form the definition of $T(x)$ in Equation (3.11)), our problem reduces to proving that $2E(x)-(1-x)K(x)$ is increasing. To this end, differentiation yields

$$[2E(x)-(1-x)K(x)]' \overset{(a)}{=} \frac{E(x)-(1-x)K(x)}{2x} \overset{(b)}{=} \frac{1}{2}T(x) \overset{(c)}{>} 0,$$

where (a) is from the differentiation identities in Lemma 3.2, (b) is from Equation (3.11), and $T(x) > 0$ follows from Lemma 3.2 (ii) together with the fact that $T(0) = 0$.

A.3.4.7 Proof of Lemma 3.14

We prove by contradiction. Suppose that $L(\hat{\alpha}; \delta) \geq F_1^{-1}(\hat{\alpha})$ at some $\hat{\alpha} \in (0, 1)$. If this is the case, then there exists a $\hat{\sigma}^2$ such that

$$F_1^{-1}(\hat{\alpha}) \leq \hat{\sigma}^2 \leq L(\hat{\alpha}; \delta). \tag{A.154}$$

Since F_1 is a decreasing function (see Lemma 3.11), the first inequality implies that $\hat{\alpha} \geq F_1(\hat{\sigma}^2)$. Then, based on the global attractiveness property in Lemma 3.9 (iii), we have

$$\psi_1(\hat{\alpha}, \hat{\sigma}^2) \leq \hat{\alpha}. \tag{A.155}$$

Further, Lemma 3.5 shows that $F_1^{-1}(\hat{\alpha}) > F_2(\hat{\alpha}; \delta)$ for $\delta > \delta_{\text{AMP}}$, and using Equation (A.154) we also have $\hat{\sigma}^2 \geq F_1^{-1}(\hat{\alpha}) > F_2(\hat{\alpha}; \delta)$. Also, from Equation (A.154),

$$\hat{\sigma}^2 \leq L(\hat{\alpha}; \delta) \overset{(a)}{<} L(0; \delta) = \frac{4}{\delta}\left(1 - \frac{\pi^2}{16}\right) < \frac{4}{\delta} \leq \sigma_{\max}^2,$$

where (a) is due to the monotonicity of $L(\alpha; \delta)$ (see Lemma 3.13). From the above discussions, $F_2(\hat{\alpha}; \delta) < \hat{\sigma}^2 < \sigma_{\max}^2$. We then have (for $\delta > \delta_{\text{AMP}}$):

$$\psi_2(\hat{\alpha}, \hat{\sigma}^2; \delta) \overset{(a)}{<} \hat{\sigma}^2 \overset{(b)}{\leq} L(\hat{\alpha}; \delta) \overset{(c)}{\leq} L\left[\psi_1(\hat{\alpha}, \hat{\sigma}^2); \delta\right], \tag{A.156}$$

where step (a) follows from the global attractiveness property in Lemma 3.10 (iv), step (b) is due to the hypothesis in Equation (A.154), step (c) is from Equation (A.155) together with the monotonicity of $L(\alpha; \delta)$ (see Lemma 3.13). Note that Equation (A.156) shows that $\psi_2(\hat{\alpha}, \hat{\sigma}^2; \delta) < L\left[\psi_1(\hat{\alpha}, \hat{\sigma}^2); \delta\right]$, which contradicts Lemma 3.12, where we proved that $\psi_2(\alpha, \sigma^2; \delta) \geq L\left[\psi_1(\alpha, \sigma^2); \delta\right]$ for any $\alpha > 0$, $\sigma^2 > 0$ and $\delta > 0$. Hence, we must have that $L(\alpha; \delta) < F_1^{-1}(\alpha)$ for any $\alpha \in (0, 1)$.

A.3.4.8 Proof of Lemma 3.15

From Equation (3.25), proving Equation (3.28) is equivalent to proving:

$$1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4\left[1 + (\phi_1^{-1}(\alpha))^2\right]} > 1 - \frac{\pi^2}{16} - \frac{1}{2}\alpha^2, \quad \forall \alpha \in (0, 1), \tag{A.157}$$

where $\phi_1 : [0,\infty) \mapsto [0,1]$ and $\phi_2 : [0,\infty) \mapsto [0,\infty)$ are defined as (see Equation (3.26a) and Equation (3.26b)):

$$\phi_1(s) = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{A.158a}$$

$$\phi_2(s) = \int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta + s^2}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta. \tag{A.158b}$$

We make a variable change:

$$\alpha = \phi_1(s).$$

Simple calculations show that Equation (A.157) can be reformulated as the following

$$\frac{1}{1+s^2}\phi_2^2(s) < \frac{\pi^2}{4} + 2\phi_1^2(s), \quad s \in (0,\infty). \tag{A.159}$$

Let us further define

$$\phi_3(s) \equiv \int_0^{\frac{\pi}{2}} (\sin^2\theta + s^2)^{\frac{1}{2}} \mathrm{d}\theta. \tag{A.160}$$

From Equation (A.158) and Equation (A.160), we have

$$\phi_2(s) = \phi_1(s) + \phi_3(s),$$

and Equation (A.159) can be reformulated as

$$[\phi_1(s) + \phi_3(s)]^2 - (1+s^2)\left[\frac{\pi^2}{4} + 2\phi_1^2(s)\right] < 0. \tag{A.161}$$

To this end, we can write the LHS of Equation (A.161) into a quadratic form of $\phi_1(s)$:

$$\begin{aligned}
&[\phi_1(s) + \phi_3(s)]^2 - (1+s^2)\left[\frac{\pi^2}{4} + 2\phi_1^2(s)\right] \\
&= \phi_1^2(s) + \phi_3^2(s) + 2\phi_1(s)\phi_3(s) - (1+s^2)\left[\frac{\pi^2}{4} + 2\phi_1^2(s)\right] \\
&= -(1+2s^2)\phi_1^2(s) + 2\phi_1(s)\phi_3(s) - \frac{\pi^2}{4}(1+s^2) + \phi_3^2(s).
\end{aligned}$$

Hence, to prove that this quadratic form is negative everywhere, it suffices to prove that the discriminant is negative, i.e.,

$$4\phi_3^2(s) + 4(1+2s^2)\left[-\frac{\pi^2}{4}(1+s^2) + \phi_3^2(s)\right] < 0,$$

or

$$\phi_3^2(s) < \frac{\pi^2}{8}(1+2s^2).$$

Finally, by Cauchy-Schwarz we have

$$\begin{aligned}\phi_3^2(s) &= \left[\int_0^{\frac{\pi}{2}} (\sin^2\theta + s^2)^{\frac{1}{2}} \mathrm{d}\theta\right]^2 \\ &\leq \int_0^{\frac{\pi}{2}} 1\mathrm{d}\theta \cdot \int_0^{\frac{\pi}{2}} \left(\sqrt{\sin^2\theta + s^2}\right)^2 \mathrm{d}\theta \\ &= \frac{\pi}{2}\left(\frac{\pi}{4} + \frac{\pi}{2}s^2\right) = \frac{\pi^2}{8}(1+2s^2),\end{aligned}$$

which completes our proof.

A.3.4.9 Proof of Lemma 3.16

From Lemma 3.10 (v), the case $\alpha > \alpha_* \approx 0.53$ is trivial since then $\psi_2(\sigma^2, \alpha; \delta_{\text{AMP}})$ is strictly increasing in $\sigma^2 \in \mathbb{R}_+$. In the rest of this proof, we assume that $\alpha < \alpha_*$. We have derived in Equation (A.112) that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} > 0 \Longleftrightarrow \alpha > \frac{1}{2\sqrt{1+s^2}} E\left(\frac{1}{1+s^2}\right) = f(s), \tag{A.162}$$

where

$$s \triangleq \frac{\sigma}{\alpha}.$$

Hence, the result of Lemma 3.16 can be reformulated as proving the following:

$$\alpha > f(s), \quad \forall s \geq \frac{\sqrt{L(\alpha; \delta_{\text{AMP}})}}{\alpha}, \qquad \alpha \in [0, \alpha_*).$$

We proceed in three steps:

(i) In Lemma 3.15, we proved that the following holds for any $\alpha \in [0, 1]$:

$$L(\alpha; \delta_{\text{AMP}}) \geq \hat{L}(\alpha, \delta_{\text{AMP}}) \triangleq \frac{4}{\delta_{\text{AMP}}}\left(1 - \frac{\pi^2}{16} - \frac{1}{2}\alpha^2\right). \tag{A.163}$$

For convenience, define

$$\hat{s}(\alpha) \triangleq \frac{\sqrt{\hat{L}(\alpha; \delta_{\text{AMP}})}}{\alpha}. \tag{A.164}$$

(ii) We prove that $f(s)$ is monotonically decreasing on $s \in [\hat{s}(\alpha), \infty)$ for $\alpha < \alpha_*$.

(iii) We prove that the following holds for $\alpha < \alpha_*$:

$$\alpha > f(\hat{s}(\alpha)).$$

Clearly, Equation (A.162) follows from the above claims. Here, we introduce the function $\hat{L}$ since $\hat{L}$ has a simple closed-form formula and is easier to manipulate than $L(\alpha)$. We next prove step (ii). From Equation (A.121), it suffices to prove that

$$\hat{s}(\alpha) > s_*, \quad \forall \alpha < \alpha_*,$$

where s_* and α_* are defined in Equation (3.25) and Equation (A.125) respectively. To this end, we note that the following holds for $\alpha < \alpha_*$:

$$\hat{s}(\alpha) = \frac{\sqrt{\hat{L}(\alpha; \delta_{\text{AMP}})}}{\alpha} > \frac{\sqrt{\hat{L}(\alpha_*; \delta_{\text{AMP}})}}{\alpha_*} \approx 1.18,$$

where the inequality follows from the fact that $\hat{L}$ in Equation (A.163) is strictly decreasing in α, and the last step is calculated from Equation (A.163) and $\alpha_* \approx 0.527$. Finally, numerical evaluation of Equation (3.25) shows that $s_* \approx 0.458$. Hence, $\hat{s}(\alpha) > s_*$, which completes the proof.

We next prove step (iii). First, simple manipulations yields

$$\hat{s}^2(\alpha) \overset{(a)}{=} \frac{\hat{L}(\alpha)}{\alpha^2} \overset{(b)}{=} \frac{4}{\delta_{\text{AMP}}} \left[\left(1 - \frac{\pi^2}{16}\right) \cdot \frac{1}{\alpha^2} - \frac{1}{2} \right], \tag{A.165}$$

where (a) is from the definition of $\hat{s}(\alpha)$ in Equation (A.164) and (b) is due to Equation (A.163). Using Equation (A.165), we further obtain

$$\alpha = \sqrt{\frac{16 - \pi^2}{4\delta_{\text{AMP}} \hat{s}^2(\alpha) + 8}}. \tag{A.166}$$

Now, from Equation (A.166) and Equation (A.119b), we have

$$\alpha - f(\hat{s}(\alpha)) > 0 \Longleftrightarrow \sqrt{\frac{16-\pi^2}{4\delta_{\text{AMP}}\hat{s}^2(\alpha)+8}} - \frac{1}{2\sqrt{1+\hat{s}^2(\alpha)}} E\left(\frac{1}{1+\hat{s}^2(\alpha)}\right) > 0. \tag{A.167}$$

We prove Equation (A.167) by showing that the following stronger result holds:

$$\sqrt{\frac{16-\pi^2}{4\delta_{\text{AMP}}t^2+8}} - \frac{1}{2\sqrt{1+t^2}} E\left(\frac{1}{1+t^2}\right) > 0, \quad \forall t \in \mathbb{R}_+. \tag{A.168}$$

For convenience, we make a variable change:

$$x \triangleq \frac{1}{1+t^2}.$$

With some straightforward calculations, we can rewrite Equation (A.168) as

$$E(x) < \sqrt{\frac{16-\pi^2}{\delta_{\text{AMP}}(1-x)+2x}}$$

The following upper bound on $E(x)$ is due to [Wang and Chu, 2013, Eqn. (1.2)]:

$$E(x) < \frac{\pi}{2}\sqrt{1-\frac{x}{2}}, \quad \forall x \in (0,1].$$

Hence, it is sufficient to prove that

$$\frac{\pi}{2}\sqrt{1-\frac{x}{2}} < \sqrt{\frac{16-\pi^2}{\delta_{\text{AMP}}(1-x)+2x}},$$

which can be reformulated as

$$\left(1-\frac{x}{2}\right)\left(\delta_{\text{AMP}} - (\delta_{\text{AMP}}-2)x\right) < \frac{4}{\pi^2}(16-\pi^2) = \delta_{\text{AMP}}$$

where the second equality follows from the definition $\delta_{\text{AMP}} = \frac{64}{\pi^2} - 4$. The above inequality holds since $0 < 1 - \frac{x}{2} < 1$ and $0 < \delta_{\text{AMP}} - (\delta_{\text{AMP}} - 2)x < \delta_{\text{AMP}}$. This completes the proof.

A.3.4.10 Proof of Lemma 3.17

From the definition of $L(\alpha;\delta)$ in Equation (3.25), we can write

$$\psi_2\left(\alpha, L(\alpha;\delta);\delta\right) = \psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right),$$

where (note that $\bar{\sigma}$ is not the conjugate of σ)

$$\bar{\sigma}^2 \triangleq 4\left(1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4\left[1+(\phi_1^{-1}(\alpha))^2\right]}\right).$$

A key observation here is that $\bar{\sigma}^2$ does not depend on δ. Clearly, Lemma 3.17 is implied by the following stronger result:

$$\frac{\partial\psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right)}{\partial\delta} < 0, \quad \forall\bar{\sigma}^2>0, \alpha>0, \delta>0,$$

which we will prove in the sequel. For convenience, we define

$$\bar{s} \triangleq \frac{\bar{\sigma}}{\alpha},\ \gamma \triangleq \frac{1}{\delta} \text{ and } s = \sqrt{\gamma}\bar{s}. \tag{A.169}$$

Using these new variables, we have

$$\begin{aligned}\psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right) &= \psi_2\left(\alpha, \gamma\bar{\sigma}^2;\gamma^{-1}\right)\\ &= 4\gamma\left((1+\gamma\bar{s}^2)\alpha^2 + 1 - \alpha\int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta+\gamma\bar{s}^2}{\left(\sin^2\theta+\gamma\bar{s}^2\right)^{\frac{1}{2}}}\mathrm{d}\theta\right),\end{aligned}$$

where the last equality is from the definition of ψ_2 in Equation (3.6b). It remains to prove that $\psi_2\left(\alpha, \gamma\bar{\sigma}^2;\gamma^{-1}\right)$ is an increasing function of γ. The partial derivative of

$\psi_2(\alpha, \sigma^2; \delta)$ w.r.t. γ is given by

$$\begin{aligned}
&\frac{\partial \psi_2\left(\alpha, \gamma\bar{\sigma}^2; \gamma^{-1}\right)}{\partial \gamma} \\
&= 4(1+2\gamma\bar{s}^2)\alpha^2 - 4\alpha\left(\int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta + \gamma\bar{s}^2}{(\sin^2\theta + \gamma\bar{s}^2)^{\frac{1}{2}}}\mathrm{d}\theta + \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\gamma^2\bar{s}^4}{(\sin^2\theta + \gamma\bar{s}^2)^{\frac{3}{2}}}\mathrm{d}\theta\right) + 4 \\
&\overset{(a)}{=} (1+2s^2)\alpha^2 - 4\alpha\left(\int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta + s^2}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}}\mathrm{d}\theta + \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{s^4}{(\sin^2\theta + s^2)^{\frac{3}{2}}}\mathrm{d}\theta\right) + 4 \\
&\overset{(b)}{=} 4(1+2s^2)\alpha^2 - 4\alpha\left(\frac{(5s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right)}{2\sqrt{1+s^2}}\right) + 4, \qquad \text{(A.170)}
\end{aligned}$$

where in step (a) we used the relationship $s^2 = \gamma\bar{s}^2$ (see Equation (A.169)), and step (b) is from the identities in Equation (A.107). From Equation (A.170), we see that $\frac{\partial\psi_2\left(\alpha,\gamma\bar{\sigma}^2;\gamma^{-1}\right)}{\partial\gamma}$ is a quadratic function of α. Therefore, to prove $\frac{\partial\psi_2\left(\alpha,\gamma\bar{\sigma}^2;\gamma^{-1}\right)}{\partial\gamma} > 0$, it suffices to show that the discriminant is negative:

$$\left(\frac{(5s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right)}{2\sqrt{1+s^2}}\right)^2 - 4(1+2s^2) < 0. \qquad \text{(A.171)}$$

Further, to prove Equation (A.171), it is sufficient to prove that the following two inequalities hold:

$$(5s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right) > 0, \qquad \text{(A.172a)}$$

and

$$(5s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right) < 4\sqrt{1+s^2}\sqrt{1+2s^2}. \qquad \text{(A.172b)}$$

We first prove Equation (A.172a). It is sufficient to prove the following

$$(4s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right) > 0. \qquad \text{(A.173)}$$

Applying a variable change $x = \frac{1}{1+s^2}$, we can rewrite Equation (A.173) as

$$\frac{4E(x) - 2(1-x)K(x)}{x} > 0.$$

The above inequality holds since

$$4E(x) - 2(1-x)K(x) > 2E(x) - 2(1-x)K(x) = 2T(x) > 0,$$

where the last equality is from the definition of $T(x)$ in Equation (3.11).

We next prove Equation (A.172b). Again, applying the variable change $x = \frac{1}{1+s^2}$ and after some straightforward manipulations, we can rewrite Equation (A.172b) as

$$h(x)/x < 0, \quad x \in (0,1),$$

where

$$h(x) \triangleq (5-x)E(x) - 2(1-x)K(x) - 4\sqrt{2-x} < 0.$$

Hence, we only need to prove $h(x) < 0$ for $0 < x < 1$. First, we note that $\lim_{x\to 1^-} h(x) = 0$, from the fact that $E(1) = 1$ and $\lim_{x\to 1^-}(1-x)K(x) = 0$ (see Lemma 3.2 (i)). We finish the proof by showing that $h(x)$ is strictly increasing in $x \in (0,1)$. Using the identities in Equation (3.13), we can obtain

$$h'(x) = \frac{3}{2}\frac{(1-x)(E(x)-K(x))}{x} + \frac{2}{\sqrt{2-x}}.$$

To prove $h'(x) > 0$, it is equivalent to prove

$$\begin{aligned}\frac{4x}{3(1-x)\sqrt{2-x}} &> K(x) - E(x) \\ &= \int_0^{\frac{\pi}{2}} \frac{1}{(1-x\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta - \int_0^{\frac{\pi}{2}} (1-x\sin^2\theta)^{\frac{1}{2}}\mathrm{d}\theta \\ &= \int_0^{\frac{\pi}{2}} \frac{x\sin^2\theta}{(1-x\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta.\end{aligned} \tag{A.174}$$

Noting $0 < x < 1$, we can get the following

$$\int_0^{\frac{\pi}{2}} \frac{x\sin^2\theta}{(1-x\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta < \int_0^{\frac{\pi}{2}} \frac{x\sin^2\theta}{1-x\sin^2\theta}\mathrm{d}\theta = \frac{\pi}{2}\left(\frac{1}{\sqrt{1-x}} - 1\right).$$

Hence, to prove Equation (A.174), it suffices to prove

$$\frac{4x}{3(1-x)\sqrt{2-x}} > \frac{\pi}{2}\left(\frac{1}{\sqrt{1-x}} - 1\right),$$

which can be reformulated as

$$\frac{8}{3\pi}\frac{1}{\sqrt{2-x}} > \frac{\sqrt{1-x}}{1+\sqrt{1-x}}.$$

The inequality holds since

$$\frac{8}{3\pi}\frac{1}{\sqrt{2-x}} > \frac{8}{3\pi}\frac{1}{\sqrt{2}} > \frac{1}{2}, \quad \forall x \in (0,1),$$

and

$$\frac{\sqrt{1-x}}{1+\sqrt{1-x}} < \frac{1}{2}, \quad \forall x \in (0,1).$$

A.3.5 Proof of Lemma 3.19

In Lemma 3.5, we proved that $F_1^{-1}(\alpha) > F_2(\alpha;\delta)$ holds for all $\alpha \in (0,1)$ when $\delta > \delta_{\text{AMP}} \approx 2.5$. Here, we will prove that $F_1^{-1}(\alpha) > F_2(\alpha;\delta)$ holds for α close to 1 when $\delta > \delta_{\text{global}} = 2$. Similar to the manipulations given in Appendix A.3.4.4, the Equation (3.56) can be re-parameterized into the following:

$$\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{(1-\gamma s^2)\sin^2\theta + s^2}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta > 1, \quad \forall s \in (0,\xi), \tag{A.176}$$

where $\gamma \triangleq 1 - \delta/4$ and $\xi = \phi_1^{-1}(\epsilon)$ (see Equation (A.134) for the definition of ϕ_1). Again, it is more convenient to express Equation (A.176) using elliptic integrals (cf. Equation (A.145))

$$\frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2} > 1, \quad \forall x \in \left(\frac{1}{1+\xi}, 1\right), \tag{A.177}$$

where we made a variable change $x \triangleq 1/(1+s^2)$. To this end, we can verify that

$$\lim_{x\to 1} \frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2} = 1.$$

To complete the proof, we only need to show that the derivative of the LHS of Equation (A.177) in a small neighborhood of $x = 1$ is strictly negative when $\delta >$

$\delta_{\text{global}} = 2$. Using the formulas listed in Section 3.1.4, we can derive the following:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2}\right)\Big|_{x\to 1} \\
=&\frac{1}{2x^3}\left(2\gamma(x-4)E(x)\cdot(1-x)K(x) + [4\gamma(1-x)+x]\cdot(1-x)K^2(x)\right. \\
&\left.+[2\gamma(2-x)-x]E^2(x)\right)\Big|_{x\to 1} \\
=&\gamma - \frac{1}{2},
\end{aligned}$$

where the last step is due to the facts that $E(x) = 1$ and $\lim_{x\to 1}(1-x)K(x) = 0$. See Section 3.1.4 for more details. Hence, the above derivative is negative if $\gamma < \frac{1}{2}$ or $\delta > 2$ by noting the definition $\gamma = 1 - \delta/4$.

A.3.6 Continuity of the partial derivative $\frac{\partial\psi_2(\alpha,\sigma^2)}{\partial\sigma^2}$ at $(\alpha, \sigma^2) = (1, 0)$

Note that in the proof of Lemma 3.10-(i) we showed that the $\lim_{(\alpha,\sigma^2)\to(1,0)} \frac{\partial\psi_2(\alpha,\sigma^2)}{\partial\sigma^2} = \frac{2}{\delta}$. Our goal here is to show that the derivative exists at $(\alpha, \sigma^2) = (1, 0)$ and it is equal to $\frac{2}{\delta}$.

A.3.6.1 Proof of the main claim

Our goal in this section is to show that $\frac{\partial\psi_2(\alpha,\sigma^2)}{\partial\sigma^2}\Big|_{(1,0)} = \frac{2}{\delta}$. From the definition of the partial derivative, we have

$$\begin{aligned}
\frac{\partial\psi_2(\alpha,\sigma^2)}{\partial\sigma^2}\Big|_{(1,0)} &= \lim_{\sigma^2\to 0}\frac{1}{\sigma^2}(\psi_2(1,\sigma^2) - \psi_2(1,0)) \\
&= \lim_{\sigma^2\to 0}\frac{4}{\delta\sigma^2}\left(1+\sigma^2+1 - \int_0^{\pi/2}\frac{2\sin^2\theta+\sigma^2}{(\sin^2\theta+\sigma^2)^{\frac{1}{2}}}d\theta - 2 + \int_0^{\pi/2}2\sin\theta d\theta\right) \\
&= \lim_{\sigma^2\to 0}\frac{4}{\delta\sigma^2}\left(\sigma^2 - \int_0^{\pi/2}\frac{2\sin^2\theta+\sigma^2}{(\sin^2\theta+\sigma^2)^{\frac{1}{2}}}d\theta + 2\right) \qquad \text{(A.178)}
\end{aligned}$$

Define $m \triangleq 1/\sigma^2$. Then,

$$\begin{aligned}
&\left.\frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2}\right|_{(1,0)} \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\Big(\frac{1}{m} - \int_0^{\pi/2} \frac{2\sqrt{m}\sin^2\theta + 1/\sqrt{m}}{(m\sin^2\theta+1)^{\frac{1}{2}}} d\theta + 2\Big) \\
&\overset{(a)}{=} \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)E(\frac{m}{m+1}) - K(\frac{m}{m+1})}{\sqrt{m(m+1)}} - \frac{1}{\sqrt{m(m+1)}} K\left(\frac{m}{m+1}\right) + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)E(\frac{m}{m+1})}{\sqrt{m(m+1)}} + \frac{1}{\sqrt{m(m+1)}} K\left(\frac{m}{m+1}\right) + 2\right). \qquad \text{(A.179)}
\end{aligned}$$

To obtain Equality (a) we have used Equation (A.107). By employing Lemma 3.2 (i) we have

$$\begin{aligned}
&\lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)E(\frac{m}{m+1})}{\sqrt{m(m+1)}} + \frac{1}{\sqrt{m(m+1)}} K\left(\frac{m}{m+1}\right) + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}}\left(1 + \frac{1}{2}\frac{\log 4\sqrt{m+1}}{m+1} - \frac{1}{4(m+1)}\right)\right. \qquad \text{(A.180)} \\
&\quad \left. + \frac{1}{\sqrt{m(m+1)}}\log 4\sqrt{m+1} + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}}\left(1 - \frac{1}{4(m+1)}\right) + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}} + 2 + \frac{1}{2\sqrt{m(m+1)}}\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}} + 2\right) + \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{2\sqrt{m(m+1)}}\right) = 0 + \frac{2}{\delta}.
\end{aligned}$$

Again we emphasize that we have also shown in the proof of Lemma 3.10 that $\lim_{(\alpha,\sigma^2)\to(1,0)} \frac{\partial\psi_2(\alpha,\sigma^2)}{\partial\sigma^2} = \frac{2}{\delta}$. Hence, $\frac{\partial\psi_2(\alpha,\sigma^2)}{\partial\sigma^2}$ is continuous at $(\alpha,\sigma^2) = (1,0)$.

A.3.7 Proof of Lemma 3.21

In the following, we will prove that each part of the lemma holds when σ_w^2 is smaller than a constant. Hence, the statements hold simultaneously when σ_w^2 is smaller than the minimum of those constants.

Part (a): In Lemma 3.10-(iii) we proved that, for the noiseless setting, $\psi_2(\alpha;\sigma^2;\delta) \leq \sigma^2_{\max}$ for $\sigma^2 \in [0, \sigma^2_{\max}]$. If fact, it is easy to verify that our proof can be strengthened to $\psi_2(\alpha;\sigma^2;\delta) \leq \sigma^2_{\max}$ for $\sigma^2 \in [0, 2]$, see Equation (A.123). Note that $\sigma^2_{\max} = \max\{1, 4/\delta\} \leq 4/\delta_{\text{AMP}} \approx 1.6$. Hence, $\psi_2(\alpha;\sigma^2;\delta) \leq \sigma^2_{\max}$ for $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}] = \sigma^2_{\max} + 4\sigma^2_w$ when σ^2_w is small. Further, $\psi_2(\alpha;\sigma^2;\delta,\sigma^2_w) = \psi_2(\alpha;\sigma^2;\delta) + 4\sigma^2_w$, and hence $\psi_2(\alpha;\sigma^2;\delta,\sigma^2_w) \leq \tilde{\sigma}^2_{\max}$ for $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$.

Part (b): The claim is a consequence of three facts: (i) $\psi_2(\alpha,\sigma^2;\delta,\sigma^2_w) \leq \sigma^2$ at $\sigma^2 = \tilde{\sigma}^2_{\max}$; (ii) $\frac{\partial\psi_2(\alpha,\sigma^2;\delta,\sigma^2_w)}{\partial\sigma^2} < 1$ when $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$, and (iii) if $\alpha \geq \alpha_*$, then $\frac{\partial\psi_2(\alpha,\sigma^2;\delta,\sigma^2_w)}{\partial\sigma^2} > 0$ for any $\sigma^2 \geq 0$. Fact (i) has been proved in part (a) of this lemma. For Fact (ii), recall that in Equation (A.124) we have proved $\frac{\partial\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2} < 1$ when $\sigma^2 \in [0, \sigma^2_{\max}]$. Again, similar to part (a) of this lemma, we can argue that the result actually holds for $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$. We prove Fact (ii) by further noting $\psi_2(\alpha,\sigma^2;\delta,\sigma^2_w) = \psi_2(\alpha,\sigma^2;\delta) + 4\sigma^2_w$ and hence $\frac{\partial\psi_2(\alpha,\sigma^2;\delta,\sigma^2_w)}{\partial\sigma^2} = \frac{\partial\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2}$. Fact (iii) follows from Lemma 3.10-(v) and the fact that $\frac{\partial\psi_2(\alpha,\sigma^2;\delta,\sigma^2_w)}{\partial\sigma^2} = \frac{\partial\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2}$.

We now show that $F_2(\alpha;\delta,\sigma^2_w)$ is a continuous function of σ^2_w. Let x be an arbitrary constant in $(0,\epsilon)$. Suppose that $\lim_{\sigma^2_w\to x^-} F_2(\alpha;\delta,\sigma^2_w) = y_1$ and $\lim_{\sigma^2_w\to x^+} F_2(\alpha;\delta,\sigma^2_w) = y_2$, where $y_1, y_2 \in [0, \tilde{\sigma}^2_{\max}]$ and $y_1 \neq y_2$. Since F_2 is the fixed point of ψ_2, we then have $y_1 = \psi_2(\alpha, y_1;\delta) + 4x$ and $y_2 = \psi_2(\alpha, y_2;\delta) + 4x$, which leads to $y_1 - \psi_2(\alpha, y_1;\delta) = y_2 - \psi_2(\alpha, y_2;\delta)$. However, we have shown in Lemma 3.10 that $\tilde{\Psi}_2(\alpha,\sigma^2;\delta) \triangleq \sigma^2 - \psi_2(\alpha,\sigma^2;\delta) - C$ is a strictly increasing function of σ^2 in $[0, \tilde{\sigma}^2_{\max}]$, and hence for any $C \in \mathbb{R}$ there cannot be two solutions to $\tilde{\Psi}_2(\alpha,\sigma^2;\delta) = 0$. This leads to contradiction.

Part (c): It is more convenient to introduce a variable change:

$$s \triangleq \phi_1^{-1}(\alpha) \quad \text{and} \quad s_\star(\delta,\sigma^2_w) = \phi_1^{-1}\left(\alpha_\star(\delta,\sigma^2_w)\right).$$

As have been argued in Appendix A.3.4.4, $F_1^{-1}(\alpha) \leq F_1^{-1}(0) = \pi^2/16 < \tilde{\sigma}^2_{\max}$. Then, by the global attractiveness of $F_2(\alpha;\delta,\sigma^2_w)$ (part (b) of this lemma) and noting that $\phi_1 : [0,\infty] \mapsto [0,1]$ is a decreasing function, our claim can be equivalently refomulated as

$$\psi_2\left(\phi_1(s), s^2\phi_1^2(s);\delta\right) + 4\sigma^2_w > s^2\phi_1^2, \quad \forall s \in [0, s_\star(\delta,\sigma^2_w)), \tag{A.181}$$

and

$$\psi_2\left(\phi_1(s), s^2\phi_1^2(s);\delta\right) + 4\sigma^2_w < s^2\phi_1^2, \quad \forall s > s_\star(\delta,\sigma^2_w).$$

From the definition of ψ_2 in Equation (3.61) and after straightforward manipulations,

we can write Equation (A.181) into

$$T(s^2,\delta,\sigma_w^2)<0,\quad \forall s\in[0,s_\star(\delta,\sigma_w^2))\quad\text{and}\quad T(s^2,\delta,\sigma_w^2)>0,\quad \forall s>s_\star(\delta,\sigma_w^2), \tag{A.182}$$

where

$$T(s^2,\delta,\sigma_w^2)\triangleq\left(1-\frac{4}{\delta}\right)\phi_1^2(s)s^2+\frac{4}{\delta}\phi_1(s)\phi_3(s)-\left(\frac{4}{\delta}+4\sigma_w^2\right). \tag{A.183}$$

From Equation (A.182), we have

$$\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2}=\left(1-\frac{4}{\delta}\right)\left(\phi_1^2(s)+2\phi_1(s)\frac{\mathrm{d}\phi_1(s)}{\mathrm{d}s^2}s^2\right)+\frac{4}{\delta}\frac{\mathrm{d}\phi_1(s)\phi_3(s)}{\mathrm{d}s^2}. \tag{A.184}$$

Applying the identities listed in Equation (3.63), we obtain

$$\left.\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2}\right|_{s=0}=1-\frac{2}{\delta}>0.$$

Further, $\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2}$ is a continuous function at $s^2=0$, and thus there exists $\epsilon>0$ such that

$$\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2}>0,\quad \forall s^2\in[0,\epsilon].$$

The above result shows that $T(s^2,\sigma_w^2)$ is monotonically increasing in $s^2\in[0,\epsilon]$. Further, from Equation (A.183) we have

$$T(s^2,\delta,\sigma_w^2)=T(s^2,\delta,0)-4\sigma_w^2.$$

It is straightforward to show that $T(0,\delta,\sigma_w^2)=-\sigma_w^2<0$. Hence, $T(s^2,\delta,\sigma_w^2)=0$ has a unique solution if the following holds:

$$\inf_{s^2\geq\epsilon}\ T(s^2,\delta,\sigma_w^2)>0,$$

or equivalently

$$4\sigma_w^2<\inf_{s^2\geq\epsilon}\ T(s^2,\delta,0). \tag{A.185}$$

Lemma 3.5 proves that $F_1^{-1}(\alpha)>F_2(\alpha;\delta)$ for $\alpha\in(0,1)$ for any $\delta>\delta_{\mathrm{AMP}}$, which, after re-parameterization implies that $T(s^2,\delta,0)>0$ for $s>0$ if $\delta>\delta_{\mathrm{AMP}}$. Hence, $\inf_{s^2\geq\epsilon}\ T(s^2,\delta,0)$ is strictly positive, and there exists sufficiently small σ_w^2 such that Equation (A.185) holds.

Part (d): From the fixed point equation $F_2 = \psi_2(\alpha, F_2; \delta, \sigma_w^2)$ where (F_2 denotes $F_2(\alpha; \delta, \sigma_w^2)$), we can derive the following (cf. Equation (A.126))

$$\left(1 - \partial_2\psi_2(\alpha, F_2; \delta, \sigma_w^2)\right) \cdot \frac{\mathrm{d}F_2(\alpha; \delta, \sigma_w^2)}{\mathrm{d}\alpha} = \partial_1\psi_2(\alpha, F_2; \delta, \sigma_w^2).$$

Similar to the proof of part (b), $1 - \partial_2\psi_2(\alpha, F_2; \delta, \sigma_w^2) > 0$ when σ_w^2 is sufficiently small. Hence, proving $\partial_1\psi_2(\alpha, F_2; \delta, \sigma_w^2) < 0$ is simplified to proving that there exists $\hat{\alpha}(\delta, \sigma_w^2)$ such that

$$\partial_1\psi_2(\alpha, F_2; \delta, \sigma_w^2) < 0, \quad \forall \alpha \in \left(0, \hat{\alpha}(\delta, \sigma_w^2)\right), \tag{A.186a}$$

and

$$\partial_1\psi_2(\alpha, F_2; \delta, \sigma_w^2) > 0, \quad \forall \alpha \in \left(\hat{\alpha}(\delta, \sigma_w^2),\, 1\right). \tag{A.186b}$$

From Equation (3.6) and after some calculations, we obtain the following

$$\begin{aligned} \frac{\partial\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2)}{\partial\alpha} &= \frac{4}{\delta}\left(2\alpha - \int_0^{\frac{\pi}{2}} \frac{2\alpha^3 \sin^4\theta + 3\alpha\sigma^2 \sin^2\theta}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{3}{2}}}\mathrm{d}\theta\right) \\ &= \frac{4}{\delta}\Bigg(2\alpha - 2\underbrace{\int_0^{\frac{\pi}{2}} \frac{\sin^4\theta + \frac{3}{2}s^2\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{3}{2}}}\mathrm{d}\theta}_{h(s)}\Bigg), \end{aligned} \tag{A.187}$$

where $s \triangleq \sigma/\alpha$. Then, we can reformulate Equation (A.186) as

$$\alpha < h\left(\frac{\sqrt{F_2(\alpha; \delta, \sigma_w^2)}}{\alpha}\right), \quad \forall \alpha \in \left(0, \hat{\alpha}(\delta, \sigma_w^2)\right),$$

and

$$\alpha > h\left(\frac{\sqrt{F_2(\alpha; \delta, \sigma_w^2)}}{\alpha}\right), \quad \forall \alpha \in \left(\hat{\alpha}(\delta, \sigma_w^2),\, 1\right).$$

From the definition given in Equation (A.187), it is easy to show that $h : \mathbb{R}_+ \mapsto [0, 1]$ is a decreasing function. Then, the above inequality can be further simplified to

$$F_2(\alpha; \delta, \sigma_w^2) < \left[\alpha \cdot h^{-1}(\alpha)\right]^2 \triangleq G(\alpha), \quad \forall \alpha \in \left(0, \hat{\alpha}(\delta, \sigma_w^2)\right), \tag{A.188a}$$

and

$$F_2(\alpha; \delta, \sigma_w^2) > \left[\alpha \cdot h^{-1}(\alpha)\right]^2 = G(\alpha), \quad \forall \alpha \in \left(\hat{\alpha}(\delta, \sigma_w^2),\, 1\right). \tag{A.188b}$$

Similar to Equation (A.181) and Equation (A.182), Equation (A.188) can be re-

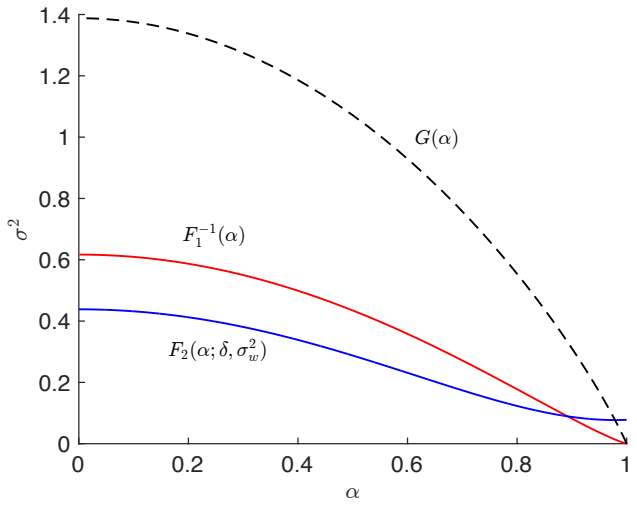

Figure A.3: Depiction of $F_1^{-1}(\alpha)$, $F_2(\alpha;\delta,\sigma_w^2)$ and $G(\alpha)$. $\alpha_\star(\delta,\sigma_w^2)$: solution to $F_1^{-1}(\alpha) = F_2(\alpha;\delta,\sigma_w^2)$. $\hat{\alpha}(\delta,\sigma_w^2)$: solution to $G^{-1}(\alpha) = F_2(\alpha;\delta,\sigma_w^2)$.

parameterized as

$$\psi_2\left(h(s), s^2\phi_1^2(s);\delta\right) + 4\sigma_w^2 > s^2h^2, \quad \forall s < \hat{s}(\delta,\sigma_w^2), \tag{A.189}$$

and

$$\psi_2\left(\phi_1(s), s^2\phi_1^2(s);\delta\right) + 4\sigma_w^2 < s^2h^2, \quad \forall s > \hat{s}(\delta,\sigma_w^2), \tag{A.190}$$

where $\hat{s}(\delta,\sigma_w^2) \triangleq h^{-1}\left(\hat{\alpha}(\delta,\sigma_w^2)\right)$. We skip the proof for Equation (A.189) since it is very similar to the proof of part (c) of this lemma. (Note that to apply the above re-parameterization (which is based on the global attractiveness of F_2, i.e., part (b) of this lemma), we need to ensure $G(\alpha) < \tilde{\sigma}_{\max}^2$. This can be seen from the fact that $G(\alpha) \le G(0) = (3\pi/8)^2 \approx 1.38$ while $\tilde{\sigma}_{\max}^2 + 4\sigma_w^2$ and $\sigma_{\max}^2 = \max\{1, 4/\delta\} > \max\{1, 4/\delta_{\text{AMP}}\} \approx 1.6$.)

Finally, to show $\hat{\alpha}(\delta,\sigma_w^2) > \alpha_\star(\delta,\sigma_w^2)$, we will prove that $G(\alpha) > F_1^{-1}(\alpha)$ for $\alpha \in [0,1)$. See the plot in Fig. A.3. Since $G(\alpha) = [\alpha \cdot h^{-1}(\alpha)]^2$ and $F_1^{-1}(\alpha) = [\alpha \cdot \phi_1^{-1}(\alpha)]^2$, we only need to prove $h^{-1}(\alpha) > \phi_1^{-1}(\alpha)$. Noting that both ϕ_1 and h are monotonically decreasing functions, it suffices to prove $h(s) > \phi_1(s)$ for $s > 0$, which directly follows

from their definitions (cf. Equation (A.187) and Equation (3.26a)):

$$\begin{aligned} h(s) - \phi_1(s) &= \int_0^{\frac{\pi}{2}} \frac{\sin^4\theta + \frac{3}{2}s^2\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{3}{2}}}\mathrm{d}\theta - \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}}\mathrm{d}\theta \\ &= \int_0^{\frac{\pi}{2}} \frac{\frac{1}{2}s^2\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{3}{2}}} > 0, \quad \forall s > 0. \end{aligned}$$

Part (e): First note that $L(\alpha;\delta,\sigma_w^2) = L(\alpha;\delta) + 4\sigma_w^2$. Hence, the proof for the claim is straightforward if the inequality $L(\alpha;\delta) < F_1^{-1}(\alpha)$ is strict for $\alpha \le \alpha_*$. This is the case since Lemma 3.14 shows that $L(\alpha;\delta) \le F_1^{-1}(\alpha)$ for $\alpha \le 1$, but equality only happends at $\alpha = 1$.

Part (f): In Lemma 3.18, we have proved the following result in the case of $\sigma_w^2 = 0$:

$$\psi_2(\alpha,\sigma^2;\delta) < F_1^{-1}(\alpha), \quad \forall 0 \le \alpha \le \alpha_*,\ L(\alpha;\delta) < \sigma^2 < F_1^{-1}(\alpha).$$

(In fact, the above inequality holds for α up to one.) In the noisy case, ψ_2 increases a little bit: $\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2) = \psi_2(\alpha,\sigma^2;\delta) + 4\sigma_w^2$. Hence, when σ_w^2 is sufficiently small, we still have

$$\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2) < F_1^{-1}(\alpha), \quad \forall 0 \le \alpha \le \alpha_*,\ L(\alpha;\delta) < \sigma^2 < F_1^{-1}(\alpha). \tag{A.191}$$

Clearly, the inequality in Equation (A.191) also holds for $L(\alpha;\delta,\sigma_w^2) < \sigma^2 < F_1^{-1}(\alpha)$, since $L(\alpha;\delta,\sigma_w^2) = L(\alpha;\delta) + 4\sigma_w^2 > L(\alpha;\delta)$.

Part (g): Note that $F_1^{-1}(\alpha_*) \approx F_1^{-1}(0.53) > 0$ does not depend on σ_w^2. Further, $F_2(1;\delta,0) = 0$ and $F_2(1;\delta,\sigma_w^2)$ is a continuous function of σ_w^2. Hence, $F_2(1;\delta,\sigma_w^2) < F_1^{-1}(\alpha_*)$ for small enough σ_w^2.

A.4 Proofs of asymptotic analysis of AMP.A in real-valued case

A.4.1 Simplifications of SE maps for Real-valued AMP.A

For the real-valued case, the SE maps are given by

$$\begin{aligned} \psi_1(\alpha,\sigma^2) &= \mathbb{E}[\partial_z g(P,Y)], \\ \psi_2(\alpha,\sigma^2;\delta,\sigma_w^2) &= \mathbb{E}\left[g^2(P,Y)\right], \end{aligned} \tag{A.192}$$

where$Z \sim \mathcal{N}(0, 1/\delta)$, $P = \alpha Z + \sigma B$ where $B \sim \mathcal{N}(0, 1/\delta)$ is independent of Z, and $Y = |Z| + W$ where $W \sim \mathcal{N}(0, \sigma_w^2)$ independent of both Z and B. Substituting $g(p, y) = y \cdot \text{sign}(p) - p$ into Equation (A.192) yields

$$\begin{aligned} \psi_1(\alpha, \sigma^2) &= \mathbb{E}\left[\partial_z |Z| \cdot \text{sign}(P)\right] \\ &= \mathbb{E}\left[\text{sign}(ZP)\right], \\ \psi_2(\alpha, \sigma^2) &= \mathbb{E}\left[\left(|Z| - |P| + W\right)^2\right] \\ &= \underbrace{\mathbb{E}\left[\left(|Z| - |P|\right)^2\right]}_{\psi_2(\alpha, \sigma^2)} + \sigma_w^2. \end{aligned} \tag{A.193}$$

Further $\mathbb{E}\left[\left(|Z| - |P|\right)^2\right] = \frac{1}{\delta}(\alpha^2 + \sigma^2 + 1) - 2\mathbb{E}[|ZP|]$. It remains to derive the following terms: $\mathbb{E}\left[\text{sign}(ZP)\right]$ and $\mathbb{E}[|ZP|]$. We first consider $\mathbb{E}\left[\text{sign}(ZP)\right]$. Similar to the derivations in Appendix Equation (A.3.1.2), we will first calculate the expectation conditioned on Z. Note that conditioned on Z, we have $P|Z \sim \mathcal{N}(\alpha Z, \sigma^2/\delta)$ and

$$S \triangleq ZP|Z \sim \mathcal{N}(\alpha Z^2, \sigma^2 Z^2/\delta). \tag{A.194}$$

We then have

$$\begin{aligned} \mathbb{E}\left[\text{sign}(ZP)\middle|Z\right] &= 2\text{Pr}(S > 0) - 1 \\ &= 2\Phi\left(\frac{\alpha}{\sigma}|Z|\sqrt{\delta}\right) - 1 \\ &= 2\Phi\left(\frac{\alpha}{\sigma}|\tilde{Z}|\right) - 1, \end{aligned}$$

where $\Phi(\cdot)$ denotes the CDF function of a standard Gaussian random variable and $\tilde{Z} \triangleq Z \cdot \sqrt{\delta} \sim \mathcal{N}(0, 1)$. We further average $\mathbb{E}_{\cdot|Z}[\text{sign}(ZP)]$ over Z:

$$\begin{aligned} \mathbb{E}\left[\text{sign}(ZP)\right] &= \mathbb{E}\left[\mathbb{E}_{\cdot|Z}[\text{sign}(ZP)]\right] \\ &= \mathbb{E}\left[2\Phi\left(\frac{\alpha}{\sigma}|\tilde{Z}|\right)\right] - 1 \\ &= \frac{2}{\pi}\text{arctan}\left(\frac{\alpha}{\sigma}\right), \end{aligned} \tag{A.195}$$

where the last step is due to the identity derived in Equation (A.96). We next derive $\mathbb{E}[|ZP|]$. Conditioned on Z, $|ZP|$ is the magnitude of a Gaussian random variable

(see Equation (A.194)), and its mean is given by Leone *et al.* [1961]

$$\begin{aligned}\mathbb{E}\left[|ZP|\middle|Z\right] &= 2\frac{\sigma|Z|}{\sqrt{\delta}}\cdot\phi\left(\frac{\alpha Z^2}{|Z|\sigma/\sqrt{\delta}}\right)+\alpha Z^2\left(1-2\Phi\left(-\frac{\alpha Z^2}{|Z|\sigma/\sqrt{\delta}}\right)\right)\\ &= 2\frac{\sigma|Z|}{\sqrt{\delta}}\cdot\phi\left(\frac{\alpha Z^2}{|Z|\sigma/\sqrt{\delta}}\right)+\alpha Z^2\left(2\Phi\left(\frac{\alpha Z^2}{|Z|\sigma/\sqrt{\delta}}\right)-1\right)\\ &= \frac{1}{\delta}\cdot\left[2\sigma\cdot|\tilde{Z}|\phi\left(\frac{\alpha|\tilde{Z}|}{\sigma}\right)+\alpha\cdot\tilde{Z}^2\left(2\Phi\left(\frac{\alpha|\tilde{Z}|}{\sigma}\right)-1\right)\right].\end{aligned}$$

Again, in the last step we defined $\tilde{Z}\triangleq\sqrt{\delta}Z$. Averaging the above equality over $|\tilde{Z}|$ yields

$$\begin{aligned}\mathbb{E}[|ZP|] &= \frac{1}{\delta}\cdot\mathbb{E}\left[2\sigma\cdot|\tilde{Z}|\phi\left(\frac{\alpha|\tilde{Z}|}{\sigma}\right)+\alpha\cdot\tilde{Z}^2\left(2\Phi\left(\frac{\alpha|\tilde{Z}|}{\sigma}\right)-1\right)\right]\\ &\overset{\text{(i)}}{=} \frac{1}{\delta}\left\{\frac{1}{\pi}\frac{2\sigma}{1+\frac{\alpha^2}{\sigma^2}}+2\alpha\cdot\left[\frac{1}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right)+\frac{1}{2}+\frac{1}{\pi}\frac{\alpha/\sigma}{1+\frac{\alpha^2}{\sigma^2}}\right]-\alpha\right\}\\ &= \frac{1}{\delta}\left\{\frac{1}{\pi}\left(\frac{2\sigma^3+2\alpha^2\sigma}{\alpha^2+\sigma^2}\right)+\frac{2\alpha}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right)\right\}\\ &= \frac{1}{\delta}\left\{\frac{2\sigma}{\pi}+\frac{2\alpha}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right)\right\},\end{aligned}\tag{A.196}$$

where Equation (i) is derived using the identities in Equation (A.96). Finally, combining Equation (A.193), Equation (A.195) and Equation (A.196), and after some calculations, we finally obtain the following

$$\begin{aligned}\psi_1(\alpha,\sigma^2) &= \frac{2}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right),\\ \psi_2(\alpha,\sigma^2;\delta,\sigma_w^2) &= \underbrace{\frac{1}{\delta}\left[\alpha^2+\sigma^2+1-\frac{4\sigma}{\pi}-\frac{4\alpha}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right)\right]}_{\psi_2(\alpha,\sigma^2;\delta)}+\sigma_w^2\end{aligned}$$

A.4.2 Proof of Theorem 3.5

The proof of Theorem 3.5 is in parallel to that for Theorem 3.2. For this reason, we will only report the discrepancies. For intuition and more discussions, please refer to Section 3.3.

A.4.2.1 Roadmap of the proof

Again, we define $F_1(\sigma^2)$ to be the non-negative fixed point of ψ_1 and $F_2(\alpha, \delta)$ to be the fixed point of ψ_2, where ψ_1 and ψ_2 are now defined in Equation (3.17). Different from the complex-valued case, ψ_2 now has a unique fixed point. Properties of ψ_1 and ψ_2 are detailed in Section A.4.2.2. Similar to complex-valued case, $F_1^{-1}(\alpha)$ and $F_2(\alpha; \delta)$ satisfy the following property:

Lemma A.13. If $\delta > \delta_{\text{AMP}} = \frac{\pi^2}{4} - 1$, then $F_1^{-1}(\alpha) > F_2(\alpha; \delta)$ for $\alpha \in (0, 1)$.

This lemma is proved in Section A.4.2.4. We will later use this lemma to show that when $\delta > \delta_{\text{AMP}} = \frac{\pi^2}{4} - 1$ the state evolution converges to the desired fixed point $(\alpha, \sigma^2) = (1, 0)$ for all initialization as long as $\alpha_0 \neq 0$. This means that AMP.A recovers the signal perfectly as long as the initial estimate is not orthogonal to the true signal.

Our next step is to analyze the dynamics of AMP.A for $\delta > \delta_{\text{AMP}}$. The following lemma implies that we only need to focus on the region where $\alpha \in [0, 1]$.

Lemma A.14. Let $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ be two sequences generated according to Equation (3.16). Then for any $\alpha_0 \geq 0$ and $\sigma_0^2 \in \mathbb{R}_+$, we have $\alpha_t \in [0, 1]$ for any $t \geq 1$.

This lemma is a direct consequence of Lemma A.19-ii proved in Section A.4.2.2. Hence, we skip its proof. Similar to Equation (3.25), the following function characterizes the lower boundary of the region that (α_t, σ_t^2) $(\forall t \geq 1)$ can fall into.

Definition 8. For any $\delta > 0$ and $\alpha \in [0, 1]$, define

$$L(\alpha; \delta) \triangleq \frac{1}{\delta}\left\{1 - \left[\frac{2}{\pi}\cos\left(\frac{\pi}{2}\alpha\right) + \alpha \sin\left(\frac{\pi}{2}\alpha\right)\right]^2\right\}. \tag{A.197}$$

For the intuition about L the reader may refer to Section 3.3. As in the complex-valued signals case, the following properties of this function play critical roles in the dynamics of the SE:

Lemma A.15. $L(\alpha; \delta)$ defined in Equation (A.197) is a strictly decreasing function of $\alpha \in (0, 1)$.

This is straightforward to see and hence the proof is skipped.

Lemma A.16. If $\delta > \delta_{\text{AMP}} = \frac{\pi^2}{4} - 1$, then $F_1^{-1}(\alpha) > L(\alpha; \delta)$ for any $\alpha \in (0, 1)$

We skip the proofs of this Lemma. The arguments are similar to Lemma 3.5 and the calculations are straightforward too. Similar to Definition 9, we divide $\{(\alpha, \sigma^2) : \alpha \in (0, 1], \sigma^2 \geq 0\}$ into four subregions.

Definition 9. We divide $\{(\alpha, \sigma^2) : \alpha \in (0, 1], \sigma^2 \geq 0\}$ into the following four sub-regions:

$$\begin{aligned}
\mathcal{R}_0 &\triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, \frac{4}{\pi^2} < \sigma^2 < \infty\right\}, \\
\mathcal{R}_1 &\triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, F_1^{-1}(\alpha) < \sigma^2 \leq \frac{4}{\pi^2}\right\}, \\
\mathcal{R}_{2a} &\triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, L(\alpha) \leq \sigma^2 \leq F_1^{-1}(\alpha)\right\}, \\
\mathcal{R}_{2b} &\triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, 0 \leq \sigma^2 < L(\alpha; \delta)\right\}.
\end{aligned} \tag{A.198}$$

Note that there are two differences between Definition 9 and Definition 6. First, the upper limit of σ^2 for $\mathcal{R}_1$ is changed from $\frac{\pi^2}{16}$ to $\frac{4}{\pi^2}$. Second, in Definition 6, $\sigma^2 < \sigma^2_{\max} = \max\{1, \delta/4\}$ for $\mathcal{R}_0$, but in Definition 9, the value of σ^2 for $\mathcal{R}_2$ is not upper bounded. Our next lemma shows that for any $(\alpha_0, \sigma_0^2) \in \mathcal{R}$, the states of the dynamical system Equation (3.16) will eventually move to $\mathcal{R}_1$ or $\mathcal{R}_{2a}$.

Lemma A.17. Suppose that $\delta > \delta_{\text{AMP}}$. Let $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ be the sequences generated according to Equation (3.16) from any $\alpha_0 > 0$ and $\sigma_0 \in \mathbb{R}_+$.

(i) Starting from $t \geq 1$, (α_t, σ_t^2) cannot be in $\mathcal{R}_{2b}$ for any $\alpha_0 \neq 0$ and $\sigma_0^2 \geq 0$.

(ii) Let (α_0, σ_0^2) be an arbitrary point in $\mathcal{R}_0$. Then, there exists a finite number $T \geq 1$ such that $(\alpha_T, \sigma_T^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$.

The proof of Lemma A.17 is very similar to that of Lemma 3.8 and therefore skipped here. Finally, we complete the proof by proving the following lemma.

Lemma A.18. Suppose that $\delta > \delta_{\text{AMP}}$. If $(\alpha_{t_0}, \sigma_{t_0}^2)$ is in $\mathcal{R}_1 \cup \mathcal{R}_{2a}$ at time t_0 (where $t_0 \geq 0$), and $\{\alpha_t\}_{t\geq t_0}$ and $\{\sigma_t^2\}_{t\geq t_0}$ are obtained via the SE in Equation (3.17), then

(i) (α_t, σ_t^2) remains in $\mathcal{R}_1 \cup \mathcal{R}_{2a}$ for all $t > t_0$;

(ii) (α_t, σ_t^2) converges:

$$\lim_{t\to\infty} \alpha_t = 1 \quad \textit{and} \quad \lim_{t\to\infty} \sigma_t^2 = 0.$$

The proof of this lemma is presented in Section A.4.2.5.

A.4.2.2 Properties of ψ_1 and ψ_2

In this section, we discuss several properties of ψ_1 and ψ_2.

Lemma A.19. $\psi_1(\alpha, \sigma^2)$ in Equation (3.17a) has the following properties (for $\alpha \geq 0$):

(i) $\psi_1(\alpha, \sigma^2)$ is a concave and strictly increasing function of $\alpha > 0$, for any given $\sigma^2 > 0$.

(ii) $0 < \psi_1(\alpha, \sigma^2) < 1$, for $\alpha > 0$ and $\sigma^2 > 0$.

(iii) If $\sigma^2 < 4/\pi^2$, then there are two nonnegative solutions to $\alpha = \psi_1(\alpha, \sigma^2)$: $\alpha = 0$ and $\alpha = F_1(\sigma^2) > 0$. Further, $F_1(\sigma^2)$ is strongly globally attracting. On the other hand, if $\sigma^2 \geq 4/\pi^2$ then $\alpha = 0$ is the unique nonnegative fixed point and it is strongly globally attracting.

Proof. The proof strategy is similar to the one given in Section 3.3.2. Also, the calculations are straightforward. Hence, to save some space we skip the proof of this lemma. □

Lemma A.20. $\psi_2(\alpha, \sigma^2; \delta)$ has the following properties:

(i) If $\delta < 1$, then $\sigma^2 = 0$ is a locally unstable fixed point to $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta)$ for any $\alpha > 0$, meaning that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}\Big|_{\sigma^2=0} > 1.$$

(ii) For any $\delta > 1$, $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta)$ has a unique fixed point, denoted as $F_2(\alpha; \delta)$, in $\sigma^2 \in [0, \infty)$ for any $\alpha \in [0, 1]$. Further, the fixed point is weakly globally attracting in $\sigma^2 \in [0, \infty)$.

(iii) For any $\delta \geq 0$, $\psi_2(\alpha, \sigma^2; \delta)$ is an increasing function of $\sigma^2 \geq 0$ if

$$\alpha > \alpha_* = \frac{1}{\pi}. \tag{A.199}$$

Further, in this case $F_2(\alpha; \delta)$ is strongly globally attracting in $\sigma^2 \in [0, \infty)$.

Proof. Recall from Equation (3.17b) that $\psi_2(\alpha, \sigma^2; \delta)$ is defined as

$$\psi_2(\alpha, \sigma^2; \delta) = \frac{1}{\delta}\left[\alpha^2 + \sigma^2 + 1 - \frac{4\sigma}{\pi} - \frac{4\alpha}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right)\right].$$

Proof of (i): The partial derivative of ψ_2 w.r.t. σ^2 is

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{1}{\delta}\left(1 - \frac{2}{\pi}\frac{\sigma}{\alpha^2 + \sigma^2}\right). \tag{A.200}$$

The claims follows from the following fact:

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}\Big|_{\sigma^2=0} = \frac{1}{\delta}, \quad \forall \alpha > 0.$$

Proof of (ii): From Equation (A.200), we see that the following holds for any $\alpha \geq 0$ and $\delta > 0$:

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} < 1, \quad \forall \sigma^2 > 0.$$

Hence, the function $\Psi_2(\alpha, \sigma^2; \delta) = \psi_2(\alpha, \sigma^2; \delta) - \sigma^2$ is strictly decreasing on $\sigma^2 \in \mathbb{R}_+$. Since $\Psi_2(\alpha, 0; \delta) = \frac{1}{\delta}(\alpha - 1)^2 \geq 0$ and $\Psi_2(\alpha, \infty; \delta) = -\infty$ for $\delta > 1$ (which is easy to show from the definition of ψ_2), it follows that there exists a unique fixed point, denoted as $F_2(\alpha; \delta)$, to the following equation:

$$\psi_2(\alpha, \sigma^2; \delta) - \sigma^2 = 0.$$

Further, using similar arguments as those in the proof of Lemma 3.10, we can prove that $F_2(\alpha; \delta)$ is globally attracting in $\sigma^2 \in [0, \infty)$.

Proof of (iii): When $\psi_2(\alpha, \sigma^2; \delta)$ is an increasing function of σ^2 in $[0, \infty)$, we have

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = 1 - \frac{2}{\pi} \frac{\sigma}{\alpha^2 + \sigma^2} > 0, \quad \forall \sigma^2 \geq 0.$$

or

$$\alpha^2 > \frac{2}{\pi}\sigma - \sigma^2, \quad \sigma^2 \geq 0.$$

It is easy to show that the maximum of the RHS over $\sigma^2 \geq 0$ is $\frac{1}{\pi^2}$. Hence, $\psi_2(\alpha, \sigma^2)$ is a strictly increasing function of σ^2 in $[0, \infty)$ if $\alpha > \frac{1}{\pi}$. □

A.4.2.3 Properties of F_1 and F_2

In this section we derive the main properties of the functions F_1 and F_2.

Lemma A.21. The following hold for $F_1(\sigma^2)$ and $F_2(\alpha; \delta)$ (for $\delta > 1$):

(i) $F_1(0) = 1$ and $\lim_{\sigma^2 \to \frac{4}{\pi^2}-} F_1(\sigma^2) = 0$. Further, by defining $F_1(\frac{4}{\pi^2}) = 0$, we have $F_1(\sigma^2)$ is continuous on $\left[0, \frac{4}{\pi^2}\right]$ and strictly decreasing in $\left(0, \frac{4}{\pi^2}\right)$;

(ii) $F_2(0; \delta) = \left(\frac{-\frac{2}{\pi} + \sqrt{\frac{4}{\pi^2} + \delta - 1}}{\delta - 1}\right)^2$ and $F_2(1; \delta) = 0$.

Proof. The proof is similar to the proof of Lemma 3.11. □

A.4.2.4 Proof of Lemma A.13

It is straightforward to show that $F_2(\alpha;\delta)$ is a decreasing function of δ for any $\alpha \in [0,1]$. Hence, we only need to prove the lemma for the case where $\delta = \delta_{\text{AMP}}$. Based on the same arguments detailed in Section A.3.4.4, it suffices to prove the following inequality:

$$\psi_2\left(\alpha, F_1^{-1}(\alpha); \delta_{\text{AMP}}\right) < F_1^{-1}(\alpha), \quad \forall \alpha \in (0,1). \tag{A.201}$$

We make the following variable change:

$$t = g^{-1}(\alpha),$$

where $g : (0,\infty) \mapsto (0,1)$ is defined as (with some abuse of notations)

$$g(t) \triangleq \frac{2}{\pi}\arctan(t). \tag{A.202}$$

Based on this re-parameterization, Equation (A.201) becomes

$$\psi_2\left(g(t), \frac{g^2(t)}{t^2}; \delta_{\text{AMP}}\right) < \frac{g^2(t)}{t^2}, \quad \forall t > 0. \tag{A.203}$$

Substituting the definition of ψ_2 in Equation (3.17b) into Equation (A.203) and after some straightforward calculations, it can be shown that Equation (A.203) is implied by the following:

$$G(t) \triangleq t^2 + \frac{4}{\pi}\frac{t}{g(t)} - \frac{t^2}{g^2(t)} > 1 - \delta_{\text{AMP}}, \quad \forall t > 0. \tag{A.204}$$

From Equation (A.202) and Equation (A.204) and noting $\delta_{\text{AMP}} = \frac{\pi^2}{4} - 1$, we can verify that $\lim_{t\to 0_+} G(t) = 1 - \delta_{\text{AMP}}$. Consequently, it suffices to prove that $G(t)$ is strictly increasing on $(0,\infty)$. To this end, we calculate $G'(t)$:

$$\begin{aligned} G'(t) &= 2t + \frac{4}{\pi}h'(t) - 2h(t)\cdot h'(t) \\ &= 2h(t)h'(t)\cdot\left(\frac{t}{h(t)h'(t)} + \frac{2}{\pi}\frac{1}{h(t)} - 1\right), \end{aligned} \tag{A.205}$$

where for convenience we defined

$$h(t) \triangleq \frac{t}{g(t)} = \frac{\pi}{2}\frac{t}{\arctan(t)}. \tag{A.206}$$

We first note that $h'(t) > 0$:

$$h'(t) = \frac{\pi}{2} \cdot \frac{\arctan(t) - \frac{t}{1+t^2}}{\arctan^2(t)} > 0, \quad t > 0, \tag{A.207}$$

where the inequality follows since $\arctan(t) - \frac{t}{1+t^2}$ is strictly increasing on $(0, \infty)$ and $[\arctan(t) - \frac{t}{1+t^2}]|_{t=0} = 0$. Hence, to prove $G'(t) > 0$, we only need to prove that (cf. Equation (A.205))

$$\underbrace{\frac{t}{h(t)h'(t)}}_{G_1(t)} + \underbrace{\frac{2}{\pi}\frac{1}{h(t)}}_{G_2(t)} - 1 > 0, \quad \forall t > 0. \tag{A.208}$$

Similar to the treatment in Appendix A.3.4.4, we consider two different cases: (1) $0 < t \leq 0.75$ and (2) $t \geq 0.75$.

(i) Case I: $0 < t \leq 0.75$. From Equation (A.207), $h(t)$ is a strictly increasing function of $t > 0$, and thus $G_2(t) = \frac{2}{\pi h(t)}$ is strictly decreasing.

We next show that $G_1(t)$ is an increasing function of $t > 0$. The derivative of $G_1(t)$ is given by:

$$\begin{aligned} G_1'(t) &\overset{(a)}{=} \left(\frac{4}{\pi^2} \frac{\arctan^3(t)}{\arctan(t) - \frac{t}{1+t^2}} \right)' \\ &= \frac{4}{\pi^2} \cdot \frac{\arctan^2(t)\left[(3+t^2)\arctan(t) - 3t\right]}{\left[t - (1+t^2)\arctan(t)\right]^2} \\ &\overset{(b)}{>} 0, \end{aligned}$$

where (a) is from Equation (A.208), Equation (A.206) and Equation (A.207), and (b) is a consequence of the following facts: (i) $[(3+t^2)\arctan(t) - 3t]_{t=0} = 0$, (ii) $[(3+t^2)\arctan(t) - 3t]' = 2t\left(\arctan(t) - \frac{t}{1+t^2}\right) > 0$ (similar to Equation (A.207)).

The following proof is based on the idea introduced in Appendix A.3.4.4: since $G_1(t)$ is an increasing function and $G_2(t)$ is a decreasing function, the following holds for any $c_2 > c_1 > 0$:

$$G_1(c_1) + G_2(c_2) - 1 > 0 \Longrightarrow G_1(t) + G_2(t) - 1 > 0, \quad \forall t \in [c_1, c_2].$$

We verified that $G_1(c_1) + G_2(c_2) - 1 > 0$ holds for a sequence of intervals:

$[c_1, c_2] = [0, 0.32]$, $[c_1, c_2] = [0.32, 0.45]$, $[c_1, c_2] = [0.45, 0.55]$, $[c_1, c_2] = [0.55, 0.64]$, $[c_1, c_2] = [0.64, 0.7]$, $[c_1, c_2] = [0.7, 0.75]$. Altogether, we proved $G_1(t) + G_2(t) - 1 > 0$ for $t \in (0, 0.75]$.

(ii) Case II: $t \geq 0.75$. From the definitions in Equation (A.208), Equation (A.206) and Equation (A.207), and based on some calculations not shown here, we write the LHS of Equation (A.208) as

$$\begin{aligned} & G_1(t) + G_2(t) - 1 \\ = \; & \frac{4}{\pi^2} \cdot \underbrace{\frac{(t^3+t) \cdot \arctan^3(t) + (t^2+1)\arctan^2(t) - \arctan(t) \cdot t}{(t^3+t)\arctan(t) - t^2}}_{R(t)} - 1. \end{aligned} \tag{A.209}$$

From $\lim_{t\to\infty} \arctan(t) = \pi/2$, it is easy to see that

$$\lim_{t\to\infty} G_1(t) + G_2(t) - 1 = 0.$$

Hence, to prove $G_1(t) + G_2(t) - 1 > 0$ for $t \geq 0.75$, it suffices to show that $R(t)$ in Equation (A.209) is strictly decreasing on $[0.75, \infty)$. To this end, we calculate $R'(t)$ below:

$$\begin{aligned} R'(t) &= \frac{(t^4-1)\arctan^3(t) + t^3 + 3(t^3+t)\arctan^2(t) - 3(t^4+t^2)\arctan(t)}{t^2(1+t^2)\left[t - (1+t^2)\arctan(t)\right]^2} \\ &\triangleq \frac{N(t)}{D(t)}. \end{aligned} \tag{A.210}$$

Since $D(t) > 0$, we have

$$R'(t) < 0 \Longleftrightarrow N(t) < 0.$$

To this end, it can be shown that

$$N'(t) = 4t^2 \cdot \arctan(t) \cdot [t \cdot \arctan^2(t) + 3 \cdot \arctan(t) - 3t].$$

Hence, to prove $N'(t) < 0$ for $t \geq 0.75$, we only need to prove

$$t \cdot \arctan^2(t) + 3 \cdot \arctan(t) - 3t < 0, \quad \forall t \geq 0.75,$$

which is equivalent to proving

$$\arctan(t) < \frac{-3+\sqrt{9+12t^2}}{2t}, \quad \forall t \geq 0.75.$$

It is proved in [Zhu, 2008, Theorem 3] that

$$\arctan(t) < \frac{8t}{3+\sqrt{25+\frac{256}{\pi^2}t^2}}, \quad \forall t > 0.$$

Hence, it suffices to prove

$$\frac{8t}{3+\sqrt{25+\frac{256}{\pi^2}t^2}} < \frac{-3+\sqrt{9+12t^2}}{2t} = \frac{6t}{3+\sqrt{9+12t^2}}, \quad \forall t \geq 0.75.$$

which, after some straightforward manipulations, reduces to

$$3+4\sqrt{9+12t^2}-3\sqrt{25+\frac{256}{\pi^2}t^2} < 0, \quad \forall t > 0.75.$$

We can verify that the above inequality holds for $t = 0.75$. We complete our proof by showing that the LHS of the above inequality is decreasing in $t \in [0.75, \infty)$:

$$\begin{aligned}\left(3+4\sqrt{9+12t^2}-3\sqrt{25+\frac{256}{\pi^2}t^2}\right)' &= 48t \cdot \left(\frac{1}{\sqrt{9+12t^2}} - \frac{1}{\sqrt{\frac{25\pi^4}{256}+\pi^2t^2}}\right) \\ &= 48t \cdot \frac{(\pi^2-12)t^2+\frac{25\pi^4}{256}-9}{T_1^2T_2+T_1T_2^2} \\ &< 0, \quad \forall t > 0.75,\end{aligned}$$

where $T_1 \triangleq \sqrt{9+12t^2}$ and $T_2 \triangleq \sqrt{\frac{25\pi^4}{256}+\pi^2t^2}$, and the last inequality can be easily proved since $(\pi^2-12)t^2+\frac{25\pi^4}{256}-9 < 0$ is a strictly decreasing function of t and $[(\pi^2-12)t^2+\frac{25\pi^4}{256}-9]_{t=0.75} < 0$.

A.4.2.5 Proof of Lemma A.18

- **Preliminaries**

Lemma A.22. For any $\alpha > 0$ and $\delta > 0$, $L(\alpha;\delta)$ satisfies

$$L(\alpha,\delta) \geq \hat{L}(\alpha;\delta) \triangleq \frac{1}{\delta}\left(1 - \frac{4}{\pi^2} - \alpha^2\right). \tag{A.211}$$

Proof. According to Definition 8 we have

$$L(\alpha;\delta) \triangleq \frac{1}{\delta}\left\{1 - \left[\frac{2}{\pi}\cos\left(\frac{\pi}{2}\alpha\right) + \alpha\sin\left(\frac{\pi}{2}\alpha\right)\right]^2\right\}. \tag{A.212}$$

Then, the inequality $\hat{L}(\alpha;\delta) \leq L(\alpha;\delta)$ is equivalent to

$$\left[\cos\left(\frac{\pi}{2}\right), \sin\left(\frac{\pi}{2}\right)\right]\left[\frac{2}{\pi}, \alpha\right]^{\mathrm{T}} \leq \sqrt{\frac{4}{\pi^2} + \alpha^2},$$

which is clear from the Cauchy-Schwartz Inequality. □

Lemma A.23. For any $\alpha \in [0,1]$, $\psi_2(\alpha, \sigma^2; \delta_{\mathrm{AMP}})$ in Equation (3.17) is an increasing function of σ^2 in $\sigma^2 \in [L(\alpha;\delta_{\mathrm{AMP}}), \infty)$.

Proof. In Lemma A.20, we proved that ψ_2 is strictly increasing on $\sigma^2 > 0$ for $\alpha \geq 1/\pi$. Hence, we only need to consider the case $\alpha < \alpha_* = \frac{1}{\pi}$. From the expression of ψ_2 in Equation (3.17), it is straightforward to see that ψ_2 is increasing on $\sigma^2 \in [\sigma_2^2(\alpha), \infty)$ (for $\alpha < 1/\pi$), where

$$\sigma_2^2(\alpha) \triangleq \left(\frac{1}{\pi} + \sqrt{\frac{1}{\pi^2} - \alpha^2}\right)^2.$$

Lemma A.22 shows that $\hat{L}(\alpha;\delta)$ is a lower bound of $L(\alpha;\delta)$ for any $\alpha \in (0,1)$. Hence, it suffices to prove that

$$\hat{L}(\alpha;\delta_{\mathrm{AMP}}) = \frac{1}{\delta_{\mathrm{AMP}}}\left(1 - \frac{4}{\pi^2} - \alpha^2\right) \geq \sigma_2^2(\alpha), \quad \alpha \in \left[0, \pi^{-1}\right]. \tag{A.213}$$

Noting $\delta_{\mathrm{AMP}} = \frac{\pi^2}{4} - 1$, it can be shown that to prove Equation (A.213) it suffices to prove

$$\frac{1}{\pi} + \frac{\pi}{2}\frac{\pi^2 - 8}{\pi^2 - 4}\alpha^2 \geq \sqrt{\frac{1}{\pi^2} - \alpha^2}, \quad \alpha \in [0, \pi^{-1}].$$

which holds since the LHS is lower bounded by $1/\pi$ while the RHS is upper bounded by $1/\pi$. □

Lemma A.24. $\psi_2(\alpha, L(\alpha,\delta);\delta)$ is a decreasing function of $\delta > 0$ for any $\alpha > 0$.

Proof. Note that we can represent $L(\alpha,\delta)$ as $\frac{1}{\delta}\bar{\sigma}^2$, where $\bar{\sigma}^2$ is a number that does not depend on δ. Hence, we will prove that $\psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right)$ is a decreasing function of δ for any fixed $\alpha > 0$ and $\bar{\sigma}^2 > 0$. From the definition of ψ_2 in Equation (3.17b), we have

$$\begin{aligned}\psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right) &= \frac{1}{\delta}\left[\alpha^2 + \frac{1}{\delta}\bar{\sigma}^2 + 1 - \frac{4\bar{\sigma}}{\pi\sqrt{\delta}} - \frac{4\alpha}{\pi}\arctan\left(\frac{\alpha\sqrt{\delta}}{\bar{\sigma}}\right)\right] \\ &\overset{(a)}{=} \frac{1}{\delta}\left[(\alpha-1)^2 + \frac{1}{\delta}\bar{\sigma}^2 - \frac{4\bar{\sigma}}{\pi\sqrt{\delta}} + \frac{4\alpha}{\pi}\arctan\left(\frac{\bar{\sigma}}{\alpha\sqrt{\delta}}\right)\right] \\ &\overset{(b)}{=} (\alpha-1)^2\beta^2 + \alpha^2\bar{s}^2\beta^4 - \frac{4\bar{s}\alpha}{\pi}\beta^3 + \frac{4\alpha}{\pi}\arctan(\beta\bar{s})\,\beta^2,\end{aligned}$$

where (a) follows from the identity $\arctan\left(\frac{1}{s}\right) = \frac{\pi}{2} - \arctan(s)$, and in (b) we introduced the following definitions:

$$\beta \triangleq \frac{1}{\sqrt{\delta}} \quad \text{and} \quad \bar{s} \triangleq \frac{\bar{\sigma}}{\alpha}.$$

We then calculate the derivative of $\psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right) = \psi_2\left(\alpha, \beta^2\bar{\sigma}^2;\beta^{-2}\right)$ w.r.t. β:

$$\begin{aligned}&\frac{\partial\psi_2\left(\alpha,\beta^2\bar{\sigma}^2;\beta^{-2}\right)}{\partial\beta} \\ &= \beta\left[2(\alpha-1)^2 + 4\alpha^2\bar{s}^2\beta^2 - \frac{12\bar{s}\alpha}{\pi}\beta + \frac{8\alpha}{\pi}\arctan(\beta\bar{s}) + \frac{4\alpha\beta}{\pi}\frac{\bar{s}}{1+\beta^2\bar{s}^2}\right] \\ &= 2\beta\left[(\alpha-1)^2 + 2\alpha^2 s^2 - \frac{6s\alpha}{\pi} + \frac{4\alpha}{\pi}\arctan(s) + \frac{2\alpha}{\pi}\frac{s}{1+s^2}\right],\end{aligned}$$

where in the last step we defined $s \triangleq \beta\bar{s}$. It suffices to prove that

$$(\alpha-1)^2 + 2\alpha^2 s^2 - \frac{6s\alpha}{\pi} + \frac{4\alpha}{\pi}\arctan(s) + \frac{2\alpha}{\pi}\frac{s}{1+s^2} > 0,$$

or

$$(1+2s^2)\alpha^2 + \left[\frac{4}{\pi}\arctan(s) + \frac{2s}{\pi(1+s^2)} - \frac{6s}{\pi} - 2\right]\alpha + 1 > 0.$$

We prove by showing that the discriminant of the above quadratic function (of

α) is negative:

$$\left[2+\frac{6s}{\pi}-\frac{4}{\pi}\arctan(s)-\frac{2s}{\pi(1+s^2)}\right]^2-4(1+2s^2)<0.$$

We next prove that the following two inequalities hold:

$$2+\frac{6s}{\pi}-\frac{4}{\pi}\arctan(s)-\frac{2s}{\pi(1+s^2)}>0, \tag{A.214}$$

and

$$2+\frac{6s}{\pi}-\frac{4}{\pi}\arctan(s)-\frac{2s}{\pi(1+s^2)}-2\sqrt{1+2s^2}<0. \tag{A.215}$$

First, Equation (A.214) follows from the following facts: (i) $2>\frac{4}{\pi}\arctan(s)$ and (ii) $3>1/(1+s^2)$. We rewrite Equation (A.215) as

$$\frac{3s}{\pi}-\frac{2}{\pi}\arctan(s)-\frac{s}{\pi(1+s^2)}<\sqrt{1+2s^2}-1=\frac{2s^2}{1+\sqrt{1+2s^2}}. \tag{A.216}$$

Using $\arctan(s)>s/(1+s^2)$ (see Equation (A.207)), we can upper bound the LHS by

$$\frac{3s}{\pi}-\frac{2}{\pi}\arctan(s)-\frac{s}{\pi(1+s^2)}<\frac{3}{\pi}\frac{s^3}{1+s^2}.$$

Hence, to prove Equation (A.216), it is sufficient to prove

$$\frac{3}{\pi}\frac{s^3}{1+s^2}<\frac{2s^2}{1+\sqrt{1+2s^2}},$$

or

$$\frac{3}{\pi}\frac{s}{1+s^2}<\frac{2}{1+\sqrt{1+2s^2}},$$

which holds since (i) LHS is an increasing function of s while the RHS is a decreasing function, and (ii) equality holds when $s\to\infty$.

□

Lemma A.25. For any $(\alpha,\sigma^2)\in\mathcal{R}_{2a}$ and $\delta\geq\delta_{\text{AMP}}=\frac{\pi^2}{4}-1$, we have $\psi_2(\alpha,\sigma^2;\delta)<F_1^{-1}(\alpha)$, where $\mathcal{R}_{2a}$ is defined in Equation (A.198).

Proof. The proof is similar to that of Lemma 3.18. We consider three different cases:

(i) $\alpha \in [\pi^{-1}, 1]$ and $\delta \in [\delta_{\text{AMP}}, \infty]$.

(ii) $\alpha \in [0, \pi^{-1})$ and $\delta \in [\delta_{\text{AMP}}, \delta_*]$.

(iii) $\alpha \in [0, \pi^{-1}]$ and $\delta \in [\delta_*, \infty)$,

where $\delta_* = \frac{1-\left[\frac{2}{\pi}\cos(0.5)+\frac{1}{\pi}\sin(0.5)\right]^2}{\frac{1}{\pi^2}} \approx 4.87$.

Case (i): In Lemma A.20, we proved that ψ_2 is strictly increasing on $\sigma^2 > 0$ for $\alpha \geq 1/\pi$. Since in $\mathcal{R}_{2a}$ $\sigma^2 < F_1^{-1}(\alpha)$, the proof of

$$\psi_2(\alpha, \sigma^2; \delta) < F_1^{-1}(\alpha)$$

on $\mathcal{R}_{2a}$ reduces to the proof of

$$\max_{\sigma^2 < F_1^{-1}(\alpha)} \psi_2(\alpha, \sigma^2; \delta) = \psi_2(\alpha, F_1^{-1}(\alpha); \delta) < F_1^{-1}(\alpha).$$

The last equality is clear from the global attractiveness of $F_2(\alpha)$ in ψ_2 that is proved in Lemma A.20-ii and the fact that $F_2(\alpha) < F_1^{-1}(\alpha)$ that is proved in Lemma A.13.

Case (ii): As shown in Equation (A.200) we have

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{1}{\delta}\left(1 - \frac{2}{\pi}\frac{\sigma}{\alpha^2 + \sigma^2}\right). \tag{A.217}$$

Hence, ψ_2 has two stationary points if $\alpha \in [0, \pi^{-1})$:

$$\sigma_1^2(\alpha) = \left(\frac{1}{\pi} - \sqrt{\frac{1}{\pi^2} - \alpha^2}\right)^2,$$
$$\sigma_2^2(\alpha) = \left(\frac{1}{\pi} + \sqrt{\frac{1}{\pi^2} - \alpha^2}\right)^2,$$

where $\sigma_1^2(\alpha)$ is a local maximum and $\sigma_2^2(\alpha)$ is a local minimum. Then, the maximum of ψ_2 over $\sigma^2 \in [L(\alpha; \delta), F_1^{-1}(\alpha)]$ can only happen at either $L(\alpha; \delta)$ or $F_1^{-1}(\alpha)$ if the following holds:

$$L(\alpha; \delta) \geq \sigma_1^2(\alpha), \quad \forall \alpha \in [0, \pi^{-1}).$$

Since $L(\alpha; \delta)$ is a decreasing function of α (which can be confirmed with a straightforward calculation of the derivative), then the following holds for $\alpha <$

π^{-1}:

$$L(\alpha;\delta) \geq L(\pi^{-1};\delta) = \frac{1}{\delta}\left\{1 - \left[\frac{2}{\pi}\cos(0.5) + \frac{1}{\pi}\sin(0.5)\right]^2\right\} \approx \frac{0.494}{\delta}.$$

Further, $\sigma_1^2(\alpha)$ is an increasing function of α and is upper bounded by

$$\sigma_1^2(\alpha) < \frac{1}{\pi^2}, \quad \forall \alpha \in [0, \pi^{-1})$$

Hence, $L(\alpha;\delta) \geq \sigma_1^2(\alpha)$ when

$$\delta < \frac{1 - \left[\frac{2}{\pi}\cos(0.5) + \frac{1}{\pi}\sin(0.5)\right]^2}{\frac{1}{\pi^2}} = \delta^* \approx 4.87.$$

Now, suppose that $\delta < \delta^*$. Then, proving that $\psi_2(\alpha, \sigma^2; \delta) < F_1^{-1}(\alpha)$ is equivalent to proving:

$$\max\{\psi_2(\alpha, L(\alpha;\delta);\delta), \psi_2(\alpha, F_1^{-1}(\alpha);\delta)\} < F_1^{-1}(\alpha).$$

The rest of the argument is similar to the ones used in the proof of Lemma 3.18. Since according to Lemma A.24 $\psi_2(\alpha, L(\alpha;\delta);\delta)$ is a decreasing function of δ, and trivially $\psi_2(\alpha, F_1^{-1}(\alpha);\delta)\}$ is a decreasing function of δ we need to prove that

$$\max\{\psi_2(\alpha, L(\alpha;\delta_{\text{AMP}});\delta_{\text{AMP}}), \psi_2(\alpha, F_1^{-1}(\alpha_{\text{AMP}});\delta_{\text{AMP}})\} \leq F_1^{-1}(\alpha). \quad \text{(A.218)}$$

Also, since according to Lemma A.23, we have

$$\begin{aligned} & \max\{\psi_2(\alpha, L(\alpha;\delta_{\text{AMP}});\delta_{\text{AMP}}), \psi_2(\alpha, F_1^{-1}(\alpha_{\text{AMP}});\delta_{\text{AMP}})\} \\ = \;& \psi_2(\alpha, F_1^{-1}(\alpha_{\text{AMP}});\delta_{\text{AMP}}) \end{aligned}$$

and Equation (A.218) simplifies to:

$$\psi_2(\alpha, F_1^{-1}(\alpha_{\text{AMP}});\delta_{\text{AMP}}) \leq F_1^{-1}(\alpha),$$

which is a simple implication of the global attractiveness of $F_2(\alpha)$ in ψ_2 that is proved in Lemma A.20-ii.

Case (iii): Since $F_1(\sigma^2)$ is the solution of $\alpha = \psi_1(\alpha, \sigma^2) = \frac{2}{\pi}\arctan(\alpha/\sigma)$, we

can show that $F_1^{-1}(\alpha) = \alpha^2 \cdot \cot^2\left(\frac{\pi}{2}\alpha\right)$. Since $F_1^{-1}(\alpha)$ is a decreasing function, we have

$$F_1^{-1}(\alpha) > F_1^{-1}(\pi^{-1}) \approx 0.339, \quad \alpha \in [0, \pi^{-1}). \tag{A.219}$$

Further, if the following holds for $\alpha \in [0, \pi^{-1})$ we would have proved that $\psi_2(\alpha, \sigma^2; \delta) < 0.25$ when $\delta > 4$:

$$\psi_2(\alpha, \sigma^2; \delta) \leq \frac{1}{\delta}, \quad \forall (\alpha, \sigma^2) \in \mathcal{R}_{2a}. \tag{A.220}$$

Noting that $F_1^{-1}(\alpha) > 0.339 > 0.25 > 1/\delta$ for $\alpha \in [0, \pi^{-1}), \delta > 4$. Comparing this result with Equation (A.219) proves that

$$\psi_2(\alpha, \sigma^2; \delta) < F_1^{-1}(\alpha), \quad \forall \alpha \in [0, \pi^{-1}), \delta > 4.$$

Finally, we prove Equation (A.220). Since $\psi_2(\alpha, \sigma^2) = \frac{1}{\delta}(\alpha^2 + \sigma^2 + 1 - \frac{4\sigma}{\pi} - \frac{4\alpha}{\pi}\text{atan}\left(\frac{\alpha}{\sigma}\right))$, we only need to prove

$$\alpha^2 + \sigma^2 - \frac{4\sigma}{\pi} - \frac{4\alpha}{\pi}\text{atan}\left(\frac{\alpha}{\sigma}\right) \leq 0, \quad \forall \alpha \in [0, \pi^{-1}), (\alpha, \sigma^2) \in \mathcal{R}_{2a}$$

which is equivalent to

$$\alpha \cdot \frac{\alpha}{\sigma} + \sigma - \frac{4}{\pi} - \frac{4}{\pi}\frac{\alpha}{\sigma}\text{atan}\left(\frac{\alpha}{\sigma}\right) \leq 0, \quad \forall \alpha \in [0, \pi^{-1}), (\alpha, \sigma^2) \in \mathcal{R}_{2a},$$

Since $\alpha < 1$, it suffices to prove

$$\frac{\alpha}{\sigma} + \sigma - \frac{4}{\pi} - \frac{4}{\pi}\frac{\alpha}{\sigma}\text{atan}\left(\frac{\alpha}{\sigma}\right) \leq 0, \quad \forall (\alpha, \sigma^2) \in \mathcal{R}_{2a}.$$

Simple differentiation shows that the maximum of the function $f(x) = x - \frac{4}{\pi}x \cdot \text{atan}(x)$ happens at x_* where $\frac{4}{\pi} \cdot \text{atan}(x_*) = 1 - \frac{4}{\pi} \cdot \frac{x_*}{1+x_*^2}$ ($x_* \approx 0.44$) and hence

$$x - \frac{4}{\pi}x \cdot \text{atan}(x) \leq x_* - \frac{4}{\pi}x_* \cdot \text{atan}(x_*) = \frac{4}{\pi}\frac{x_*^2}{1 + x_*^2} \approx \frac{4}{\pi} \cdot 0.17 < \frac{2}{\pi}.$$

Using the above inequality, we obtain

$$\frac{\alpha}{\sigma} + \sigma - \frac{4}{\pi} - \frac{4}{\pi}\frac{\alpha}{\sigma}\text{atan}\left(\frac{\alpha}{\sigma}\right) < \sigma - \frac{2}{\pi} < 0,$$

where the last inequality is due to the fact that $(\alpha, \sigma^2) \in \mathcal{R}_{2a}$ and hence $\sigma^2 \leq F_1^{-1}(0) = \left(\frac{2}{\pi}\right)^2$. □

- **Main part** The proof is similar to that of Lemma 3.7. The only noticeable difference is the proof for the following inequality (cf. Equation (3.33))

$$\psi_2(\alpha;\sigma^2) < F_1^{-1}(\alpha), \quad \forall(\alpha,\sigma^2) \in \mathcal{R}_{2a}, \tag{A.221}$$

where $\mathcal{R}_{2a}$ is now defined in Definition 9. We have dedicated Lemma A.25 to the proof of the above inequality, which is in parallel to Lemma 3.18 for the complex-valued case.

A.4.3 Proof of Theorem 3.6

The proof is similar to that of Theorem 3.3. Hence, we only focus on the discrepancies.

A.4.3.1 $\delta > \delta_{\text{global}}$

Lemma A.26. Suppose that $\delta > \delta_{\text{global}} = 1 + \frac{4}{\pi^2}$. Then, there exists an $\epsilon > 0$ such that the following holds:

$$F_1^{-1}(\alpha) > F_2(\alpha;\delta), \quad \forall\alpha \in (1-\epsilon, 1). \tag{A.222}$$

Proof. Equation (A.222) can be re-parameterized as

$$\psi_2\left(g\left(s^{-1}\right), s^2 \cdot g^2\left(s^{-1}\right); \delta\right) < s^2 \cdot g^2\left(s^{-1}\right), \quad \forall s \in (0,\xi),$$

where $g(x) \triangleq \frac{2}{\pi}\arctan(x)$, $s = \cot(\frac{\pi}{2}\alpha)$ and $\xi = \tan(\frac{\pi}{2}\epsilon)$.

$$s \cdot \arctan\left(\frac{1}{s}\right) > \underbrace{\frac{-s^2 + s\sqrt{\frac{\pi^2}{4} + \left(1 + \frac{\pi^2}{4}(\delta-1)\right)s^2}}{1 + (\delta-1)s^2}}_{R(s)}, \quad s \in (0,\xi).$$

Taylor expansions of the LHS and the RHS are respectively given by

$$\begin{aligned} s \cdot \arctan\left(\frac{1}{s}\right) &= \frac{\pi s}{2} - s^2 + \frac{s^4}{3} + O(s^6), \\ R(s) &= \frac{\pi s}{2} - s^2 + \left(\frac{1}{\pi} - \frac{\pi}{4}(\delta-1)\right)s^3 + (\delta-1)s^4 + O(s^5) \end{aligned}$$

Then,

$$\delta > 1 + \frac{4}{\pi^2} \Longrightarrow \frac{1}{\pi} - \frac{\pi}{4}(\delta-1) < 0,$$

and in this case there exists a constant $\xi > 0$ such that

$$s \cdot \arctan\left(\frac{1}{s}\right) > R(s), \quad \forall s \in (0, \xi).$$

□

Since the rest of the proof is exactly similar to the proof of Lemma 3.3 for the sake of brevity we skip it here.

A.4.3.2 $\delta < \delta_{\text{global}}$

It is straightforward to use an argument similar to the one presented in Section 3.4.2 and show that there exists a neighborhood of $(\alpha, \sigma^2) = (1, 0)$ in which $\psi_2(\alpha, \sigma^2) - \sigma^2 > 0$. Hence, the state evolution moves away from $(0, 1)$.

A.4.4 Proofs of Theorems 3.7

In the noisy setting, the state evolution of AMP.A becomes

$$\psi_1(\alpha, \sigma^2) = \frac{2}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right), \tag{A.223a}$$

$$\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) = \frac{1}{\delta}\left[\alpha^2 + \sigma^2 + 1 - \frac{4\sigma}{\pi} - \frac{4\alpha}{\pi}\arctan\left(\frac{\alpha}{\sigma}\right)\right] + \sigma_w^2. \tag{A.223b}$$

Similar to the complex-valued case, the SE of real-valued AMP.A still converges to the nonzero fixed point, as stated in Lemma A.27 below. We skip the proof since it is very similar to the proof of Lemma 3.22.

Lemma A.27. Let $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ be two state sequences generated according to Equation (3.16) from $\alpha_0 > 0$ and $\sigma_0^2 < \infty$. Then, for any $\delta > \delta_{\text{AMP}}$ the following holds for sufficiently small σ_w^2:

$$\lim_{t\to\infty} \alpha_t = \alpha_\star(\delta, \sigma_w^2) \quad \text{and} \quad \lim_{t\to\infty} \sigma_t^2 = \sigma_\star^2(\delta, \sigma_w^2),$$

where $\alpha_\star(\delta, \sigma_w^2)$ is the unique positive solution to $F_1^{-1}(\alpha) = F_2(\alpha; \delta, \sigma_w^2)$ and $\sigma_\star^2(\delta, \sigma_w^2) = F_1^{-1}(\alpha_\star(\delta, \sigma_w^2))$.

Now we can prove Theorem 3.7. First note that $\text{AMSE}(\sigma_w^2, \delta) = (\alpha - 1)^2 + \sigma^2$, where with slight abuse of notation α and σ^2 denote the solution of Equation (A.223)

(which are also functions of σ_w^2 and δ), i.e.,

$$\alpha = \frac{2}{\pi}\text{arctan}\left(\frac{\alpha}{\sigma}\right), \tag{A.224a}$$

$$\sigma^2 = \frac{1}{\delta}\left[\alpha^2 + \sigma^2 + 1 - \frac{4\sigma}{\pi} - \frac{4\alpha}{\pi}\text{arctan}\left(\frac{\alpha}{\sigma}\right)\right] + \sigma_w^2. \tag{A.224b}$$

Using Equation (A.224a) and with simple manipulations we can rewrite Equation (A.224b) as

$$(\delta - 1)\sigma^2 + \alpha^2 + \frac{4\sigma}{\pi} - 1 - \delta\sigma_w^2 = 0. \tag{A.225}$$

We make the following variable change:

$$s \triangleq \frac{\sigma}{\alpha}.$$

From Equation (A.224a) and the definition of s, we have

$$\alpha = \frac{2}{\pi}\text{arctan}(s^{-1}) \quad \text{and} \quad \sigma = \frac{2}{\pi}\text{arctan}(s^{-1}) \cdot s. \tag{A.226}$$

Substituting Equation (A.226) into Equation (A.225) yields

$$T(s^2, \sigma_w^2) \triangleq \left[(\delta - 1)s^2 + 1\right] \cdot \text{arctan}^2(s^{-1}) + 2 \cdot s \cdot \text{arctan}(s^{-1}) - \frac{\pi^2}{4}(1 + \delta\sigma_w^2) = 0. \tag{A.227}$$

We have

$$\begin{aligned}
\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2} =& \frac{1}{2s}\left(2s(\delta - 1)\text{arctan}^2\left(s^{-1}\right)\right. \\
&\left. -2\left[(\delta - 1)s^2 + 1\right]\frac{\text{arctan}\left(s^{-1}\right)}{1 + s^2} + 2\text{arctan}\left(s^{-1}\right) - \frac{2s}{1 + s^2}\right) \\
=& \ (\delta - 1) \cdot \text{arctan}^2\left(s^{-1}\right) - \text{arctan}\left(s^{-1}\right)\frac{(\delta - 1)s}{1 + s^2} + \text{arctan}\left(s^{-1}\right)\frac{s}{1 + s^2} - \frac{1}{1 + s^2}, \\
\frac{\partial T(s^2, \sigma_w^2)}{\partial \sigma_w^2} =& -\frac{\pi^2}{4}\delta.
\end{aligned}$$

Note that we have an implicit relation between s^2 and σ_w^2. By the implicit function

theorem we have

$$\begin{aligned}\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2} &= -\frac{\partial T(s^2,\sigma_w^2)}{\partial \sigma_w^2}\left(\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2}\right)^{-1} \\ &= \frac{\frac{\pi^2}{4}\delta}{(\delta-1)\cdot \arctan^2\left(s^{-1}\right) - \arctan\left(s^{-1}\right)\frac{(\delta-1)s}{1+s^2} + \arctan\left(s^{-1}\right)\frac{s}{1+s^2} - \frac{1}{1+s^2}}.\end{aligned}$$

Furthermore, from Equation (A.227), we see that $s^2 = 0$ when $\sigma_w^2 = 0$ and hence

$$\left.\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2}\right|_{\sigma_w^2=0} = \frac{\frac{\pi^2}{4}\delta}{\frac{\pi^2}{4}(\delta-1)-1} = \frac{\delta}{\delta - \left(1+\frac{4}{\pi^2}\right)},$$

where we defined $\arctan(s^{-1}) = \pi/2$ at $s = 0$. Now it is straightforward to use the mean value theorem to prove that

$$\lim_{\sigma_w^2\to 0}\frac{s^2}{\sigma_w^2} = \left.\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2}\right|_{\sigma_w^2=0} = \frac{\delta}{\delta-\left(1+\frac{4}{\pi^2}\right)}.$$

Further, notice that

$$\begin{aligned}\mathrm{AMSE}(\sigma_w^2,\delta) &= (\alpha-1)^2 + \sigma^2 \\ &= \left[\frac{2}{\pi}\arctan(s^{-1}) - 1\right]^2 + \left[\frac{2}{\pi}\arctan(s^{-1})\cdot s\right]^2,\end{aligned}$$

and it is straightforward to show that

$$\lim_{s^2\to 0}\frac{\mathrm{AMSE}(\sigma_w^2,\delta)}{s^2} = 1 + \frac{4}{\pi^2}.$$

Hence,

$$\begin{aligned}\lim_{\sigma_w^2\to 0}\frac{\mathrm{AMSE}(\sigma_w^2,\delta)}{\sigma_w^2} &= \lim_{s^2\to 0}\frac{\mathrm{AMSE}(\sigma_w^2,\delta)}{s^2}\cdot\lim_{\sigma_w^2\to 0}\frac{s^2}{\sigma_w^2} \\ &= \left(1+\frac{4}{\pi^2}\right)\cdot\frac{\delta}{\delta-\left(1+\frac{4}{\pi^2}\right)},\end{aligned}$$

which proves Theorem 3.7 by noting that $\delta_{\mathrm{global}} = 1 + 4/\pi^2$.